MTH 464/564

Lectures 12-17

Yevgeniy Kovchegov
Oregon State University

- Indicator variables.
- Conditional distributions.
- Conditional expectation.
- Wald's equation.
- Conditional variance.
- The law of total variance.
- Variance of a random sum of random variables.

- Conditional expectation as a projection.

- The law of total variance via Pythagorean Theorem.

- Randomization formulas.

- Joint cumulative distribution function.

- Moment generating functions.

Indicator variables.

• **Example.** A Binomial random variable S_n with parameters (n, p) represents the number of success in n independent Bernoulli trials, each having probability p of success and $1-p$ of failure.

Consider Bernoulli random variables

$$X_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ trial is a success,} \\ 0 & \text{if the } i^{\text{th}} \text{ trial is a failure.} \end{cases}$$

For each $i = 1, \ldots, n$, X_i is the indicator variable for the event that the i^{th} trial is a success.

$$\text{Then, } \quad S_n = X_1 + X_2 + \ldots + X_n,$$

where $E[X_i] = p$ and $Var(X_i) = p(1-p)$ for all $i = 1, \ldots, n$.

Hence,

$$E[S_n] = E[X_1 + \ldots + X_n] = E[X_1] + \ldots + E[X_n] = np$$

and

$$Var(S_n) = Var(X_1 + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n) = np(1-p).$$

Indicator variables.

- **Definition.** Consider an event A. The random variable
$$X = \begin{cases} 1 & \text{if the event } A \text{ occurs} \\ 0 & \text{if the event } A \text{ does not occur} \end{cases}$$
is said to be the indicator variable for the event A.

- Frequently used notation: I_A.

- X is a Bernoulli random variable:
$$E[X] = P(A) \quad \text{and} \quad Var(X) = P(A)\big(1 - P(A)\big)$$

- The indicator variable of the complement $\overline{A}$ of A is $I_{\overline{A}} = 1 - I_A$.
- For all $k \neq 0$ and $X = I_A$, we have $X^k = X$ and $E[X^k] = P(A)$.
- For given events A and B the indicator variables $X = I_A$ and $Y = I_B$ satisfy $XY = I_{A\cap B}$ (i.e., $I_A\, I_B = I_{A\cap B}$) and
$$E[XY] = P(A \cap B).$$

Also, by de Morgan's law, $1 - (1 - X)(1 - Y) = I_{A\cup B}$ and
$$1 - E[(1 - X)(1 - Y)] = P(A \cup B).$$

Indicator variables.

- General case of Inclusion-Exclusion Theorem.

$$P(E_1 \cup E_2 \cup \ldots \cup E_n) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1 < i_2 < \ldots < i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r})$$

$$= \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2}) + \sum_{i_1 < i_2 < i_3} P(E_{i_1} \cap E_{i_2} \cap E_{i_3}) - \ldots + (-1)^{n+1} P(E_1 \cap E_2 \cap \ldots \cap E_n)$$

- **Example.** $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.

- **Example.**

$$P(E_1 \cup E_2 \cup E_3) = \sum_{i=1}^{3} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2}) + P(E_1 \cap E_2 \cap E_3)$$

$$= P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3)$$

Indicator variables.

- General case of Inclusion-Exclusion Theorem.

$$P(E_1 \cup E_2 \cup \ldots \cup E_n) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1<i_2<\ldots<i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r})$$

$$= \sum_{i=1}^{n} P(E_i) - \sum_{i_1<i_2} P(E_{i_1} \cap E_{i_2}) + \sum_{i_1<i_2<i_3} P(E_{i_1} \cap E_{i_2} \cap E_{i_3}) - \ldots + (-1)^{n+1} P(E_1 \cap E_2 \cap \ldots \cap E_n)$$

- **Proof.** Let

$$X_i = \begin{cases} 1 & \text{if the event } E_i \text{ occurs} \\ 0 & \text{if the event } E_i \text{ does not occur} \end{cases}$$

Then,

$$X_i\,X_j = \begin{cases} 1 & \text{if the event } E_i \cap E_j \text{ occurs} \\ 0 & \text{if the event } E_i \cap E_j \text{ does not occur} \end{cases}$$

and, by de Morgan's law,

$$1-(1-X_i)\,(1-X_j) = \begin{cases} 1 & \text{if the event } E_i \cup E_j \text{ occurs} \\ 0 & \text{if the event } E_i \cup E_j \text{ does not occur} \end{cases}$$

Indicator variables.

• **Proof (cont.):** Let $X_i = \begin{cases} 1 & \text{if the event } E_i \text{ occurs} \\ 0 & \text{if the event } E_i \text{ does not occur} \end{cases}$

Then,

$$1-(1-X_1)(1-X_2)\ldots(1-X_n) = \begin{cases} 1 & \text{if the event } E_1 \cup E_2 \cup \ldots \cup E_n \text{ occurs} \\ 0 & \text{if the event } E_1 \cup E_2 \cup \ldots \cup E_n \text{ does not occur} \end{cases}$$

and

$$P(E_1 \cup E_2 \cup \ldots \cup E_n) = E\Big[1-(1-X_1)(1-X_2)\ldots(1-X_n)\Big]$$

$$= \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1<i_2<\ldots<i_r} E[X_{i_1} X_{i_2} \ldots X_{i_r}]$$

$$= \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1<i_2<\ldots<i_r} P(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_r})$$

as

$$1-(1-X_1)(1-X_2)\ldots(1-X_n) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1<i_2<\ldots<i_r} X_{i_1} X_{i_2} \ldots X_{i_r}$$

Indicator variables.

• **Example.** Consider performing independent Bernoulli trials, each with probability p of success and probability $1-p$ of failure. Recall that a geometric random variable with parameter p counts the number of trials until the first success.

Let X be a geometric random variable. We want to find $E[X]$ using indicator variables.

Let F_i denote the event of failure on the i^{th} trial, and let X_i denote its indicator variable, i.e.,

$$X_i = I_{F_i}$$

Then, $X = 1 + X_1 + X_1X_2 + X_1X_2X_3 + \ldots$ and, by independence of X_i, we have

$$\begin{aligned}
E[X] &= 1 + E[X_1] + E[X_1X_2] + E[X_1X_2X_3] + \ldots \\
&= 1 + E[X_1] + E[X_1]E[X_2] + E[X_1]E[X_2]E[X_3] + \ldots \\
&= 1 + P(F_1) + P(F_1)P(F_2) + P(F_1)P(F_2)P(F_3) + \ldots \\
&= 1 + (1-p) + (1-p)^2 + (1-p)^3 + \ldots = \frac{1}{p}.
\end{aligned}$$

Conditional distributions: discrete variables.

Definition. Suppose X and Y are discrete random variables with joint probability mass function $p(x,y)$. For a given y such that $p_{\mathsf{y}}(y) > 0$, the conditional probability mass function $p_{X|Y}(x|y)$ is defined as

$$p_{X|Y}(x|y) = P\Big(X = x \,\Big|\, Y = y\Big) = \frac{p(x,y)}{p_{\mathsf{y}}(y)} \qquad \forall x \text{ s.t. } p(x,y) > 0.$$

Properties: • If X and Y independent, $p_{X|Y}(x|y) = p_{\mathsf{x}}(x)$.

• $p_{X|Y}(x|y)$ is a probability mass function:

$$\sum_{x:\, p_{X|Y}(x|y)>0} p_{X|Y}(x|y) = \frac{1}{p_{\mathsf{y}}(y)} \sum_{x:\, p(x,y)>0} p(x,y) = \frac{p_{\mathsf{y}}(y)}{p_{\mathsf{y}}(y)} = 1.$$

• The conditional cumulative distribution function: for a given y such that $p_{\mathsf{y}}(y) > 0$,

$$F_{X|Y}(x|y) = P\Big(X \le x \,\Big|\, Y = y\Big) = \sum_{a:a\le x} p_{X|Y}(a|y)$$

is a non-decreasing function such that

$$\lim_{x\to-\infty} F_{X|Y}(x|y) = 0 \quad \text{and} \quad \lim_{x\to\infty} F_{X|Y}(x|y) = 1.$$

Conditional distributions: discrete variables.

Definition. Suppose X and Y are discrete random variables with joint probability mass function $p(x,y)$. For a given y such that $p_{\mathrm{y}}(y) > 0$, the conditional probability mass function $p_{X|Y}(x|y)$ is defined as

$$p_{X|Y}(x|y) = P\Big(X = x \,\Big|\, Y = y\Big) = \frac{p(x,y)}{p_{\mathrm{y}}(y)} \qquad \forall x \text{ s.t. } p(x,y) > 0.$$

- $p_{X|Y}(x|y)$ is a probability mass function:

$$\sum_{x:\, p_{X|Y}(x|y)>0} p_{X|Y}(x|y) = \frac{1}{p_{\mathrm{y}}(y)} \sum_{x:\, p(x,y)>0} p(x,y) = \frac{p_{\mathrm{y}}(y)}{p_{\mathrm{y}}(y)} = 1.$$

- The conditional cumulative distribution function $F_{X|Y}(x|y)$ is defined as follows:

$$F_{X|Y}(a|y) = P\Big(X \le a \,\Big|\, Y = y\Big) = \sum_{x:\, x \le a} p_{X|Y}(x|y)$$

- Conditional probability: $P\Big(X \in A \,\Big|\, Y = y\Big) = \sum_{x \in A} p_{X|Y}(x|y).$

Conditional distributions: discrete variables.

Definition. Suppose X and Y are discrete random variables with joint probability mass function $p(x,y)$. For a given y such that $p_{\mathrm{y}}(y) > 0$, the conditional probability mass function $p_{X|Y}(x|y)$ is defined as

$$p_{X|Y}(x|y) = P\Big(X = x \,\Big|\, Y = y\Big) = \frac{p(x,y)}{p_{\mathrm{y}}(y)} \qquad \forall x \text{ s.t. } p(x,y) > 0.$$

• **Example.** Let X be Poisson with parameter λ_1 and Y be Poisson with parameter λ_2. Suppose X and Y are independent. We know that $Z = X+Y$ is Poisson with parameter $\lambda_1+\lambda_2$. We want to find the conditional probability mass function $p_{X|Z}(k|n)$ for a given integer $n \geq 0$, and $k = 0, 1, \ldots, n$.

$$p_{X|Z}(k|n) = P\Big(X = k \,\Big|\, Z = n\Big) = \frac{P\Big(X = k \bigcap X+Y = n\Big)}{P(Z = n)}$$

$$= \frac{P\Big(X = k \bigcap Y = n-k\Big)}{P(Z = n)} = \frac{P(X = k)P(Y = n-k)}{P(Z = n)}$$

Conditional distributions: discrete variables.

• **Example (continued).** Let X be Poisson with parameter λ_1 and Y be Poisson with parameter λ_2. Suppose X and Y are independent. We know that $Z = X + Y$ is Poisson with parameter $\lambda_1 + \lambda_2$.

Then, for a given $n \geq 0$, the conditional probability mass function

$$p_{X|Z}(k|n) = \frac{P(X=k)P(Y=n-k)}{P(Z=n)} = \frac{e^{-\lambda_1}\frac{\lambda_1^k}{k!}\, e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!}}{e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}}$$

$$= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k} \qquad k = 0, 1, \ldots, n.$$

Thus, conditioned on $X + Y = n$, X is a binomial random variable with parameters $\left(n,\, p = \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$.

Conditional distributions: continuous variables.

Definition. Suppose X and Y are continuous random variables with joint probability density function $f(x,y)$. For a given y such that $f_{\mathrm{y}}(y) > 0$, the conditional probability density function $f_{X|Y}(x|y)$ is defined as

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_{\mathrm{y}}(y)} \qquad \forall x \in \mathbb{R}.$$

- $f_{X|Y}(x|y)$ is a probability density function:

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y)dx = \frac{1}{f_{\mathrm{y}}(y)} \int_{-\infty}^{\infty} f(x,y)dx = \frac{f_{\mathrm{y}}(y)}{f_{\mathrm{y}}(y)} = 1.$$

- The conditional cumulative distribution function $F_{X|Y}(x|y)$ is defined as follows:

$$F_{X|Y}(a|y) = P\Big(X \le a \,\Big|\, Y = y\Big) = \int_{-\infty}^{a} f_{X|Y}(x|y)dx$$

- Conditional probability: $P\Big(X \in A \,\Big|\, Y = y\Big) = \int_A f_{X|Y}(x|y)dx.$

Conditional distributions: continuous variables.

Definition. Suppose X and Y are continuous random variables with joint probability density function $f(x, y)$. For a given y such that $f_{\mathsf{y}}(y) > 0$, the conditional probability density function $f_{X|Y}(x|y)$ is defined as

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_{\mathsf{y}}(y)} \qquad \forall x \in \mathbb{R}.$$

• **Example.** Let X and Y be continuous random variables with joint probability density function

$$f(x,y) = \begin{cases} \frac{1}{y} e^{-(y^2+x)/y} & \text{if } x > 0 \text{ and } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

For a given $y > 0$, find $f_{X|Y}(x|y)$ and $F_{X|Y}(x|y)$.

Conditional distributions: continuous variables.

• **Example (continued).** Let X and Y be continuous random variables with joint probability density function

$$f(x,y) = \begin{cases} \frac{1}{y}e^{-(y^2+x)/y} & \text{if } x>0 \text{ and } y>0, \\ \\ 0 & \text{otherwise.} \end{cases}$$

For a given $y>0$, observe that $\frac{1}{y}e^{-(y^2+x)/y} = \frac{1}{y}e^{-y}e^{-x/y}$ and

$$f_{\mathsf{y}}(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = e^{-y}\int_{0}^{\infty} \frac{1}{y}e^{-x/y}\,dx = e^{-y}$$

Thus,

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_{\mathsf{y}}(y)} = \frac{1}{y}e^{-x/y}$$

Hence, conditioned on the event $Y=y$, random variable X is exponential with parameter $\frac{1}{y}$. Finally,

$$F_{X|Y}(a|y) = \int_{-\infty}^{a} f_{X|Y}(x|y)dx = \int_{0}^{a} \frac{1}{y}e^{-x/y}dx = 1 - e^{-a/y} \qquad \forall a>0.$$

Conditional expectation: discrete variables.

Definition. Suppose X and Y are discrete random variables with joint probability mass function $p(x,y)$. For a given y such that $p_{\mathsf{y}}(y) > 0$, the conditional expectation $E[X\,|Y = y]$ is defined as

$$E[X\,|Y = y] \;=\; \sum_{x:\,p(x,y)>0} x\,p_{X|Y}(x|y)$$

• **Example.** Let X be Poisson with parameter λ_1 and Y be Poisson with parameter λ_2. Suppose X and Y are independent. We know that $Z = X + Y$ is Poisson with parameter $\lambda_1 + \lambda_2$.

For a given integer $n \geq 0$, conditioned on $Z = n$, X is a binomial random variable with parameters $\left(n,\, p = \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$, i.e.,

$$p_{X|Z}(k|n) = \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k} \qquad k = 0, 1, \ldots, n.$$

Then,

$$E[X\,|Z = n] = \sum_{k=0}^{n} k\,p_{X|Z}(k|n) = np = \frac{\lambda_1 n}{\lambda_1+\lambda_2}.$$

Conditional expectation: continuous variables.

Definition. Suppose X and Y are continuous random variables with joint probability density function $f(x,y)$. For a given y such that $f_{\mathrm{y}}(y) > 0$, the conditional expectation $E[X\,|Y=y]$ is defined as

$$E[X\,|Y=y] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx$$

• **Example.** Let X and Y be continuous random variables with joint probability density function

$$f(x,y) = \begin{cases} \frac{1}{y}e^{-(y^2+x)/y} & \text{if } x > 0 \text{ and } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

For a given $y > 0$, we know that $f_{X|Y}(x|y) = \frac{1}{y}e^{-x/y}$, i.e., conditioned on the event $Y = y$, random variable X is exponential with parameter $\frac{1}{y}$. Therefore,

$$E[X\,|Y=y] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx = \int_{0}^{\infty} x\, \frac{1}{y}e^{-x/y}\, dx = y$$

Conditional expectation: random variable $E[X|Y]$

$$E[X\,|Y=y] = \begin{cases} \displaystyle\sum_{x:\,p(x,y)>0} x\, p_{X|Y}(x|y) & \text{in discrete case} \\ \displaystyle\int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx & \text{in continuous case} \end{cases}$$

Observe that in either case, $\varphi(y) = E[X\,|Y=y]$ is a function of y.

Random variable $E[X|Y]$ is defined by letting

$$E[X|Y] = \varphi(Y)$$

• **Example.** Let X be Poisson with parameter λ_1 and Y be Poisson with parameter λ_2. Suppose X and Y are independent, and let $Z = X + Y$. For a given integer $n \geq 0$, we know that $E[X\,|Z=n] = \frac{\lambda_1 n}{\lambda_1+\lambda_2}$. Therefore,

$$E[X\,|Z] = \frac{\lambda_1}{\lambda_1+\lambda_2} Z.$$

Conditional expectation: random variable $E[X|Y]$

$$E[X\,|Y=y] = \begin{cases} \displaystyle\sum_{x:\,p(x,y)>0} x\, p_{X|Y}(x|y) & \text{in discrete case} \\ \displaystyle\int_{-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx & \text{in continuous case} \end{cases}$$

Observe that in either case, $\varphi(y) = E[X\,|Y=y]$ is a function of y.

Random variable $E[X|Y]$ is defined by letting

$$E[X|Y] = \varphi(Y)$$

- **Example.** Let X and Y be continuous random variables with

$$f(x,y) = \begin{cases} \frac{1}{y} e^{-(y^2+x)/y} & \text{if } x>0 \text{ and } y>0, \\ 0 & \text{otherwise.} \end{cases}$$

For a given $y>0$, we know that $E[X\,|Y=y] = y$. Therefore,

$$E[X\,|Y] = Y.$$

Conditional expectation: random variable $E[X|Y]$

$$E[X\,|Y=y] = \begin{cases} \displaystyle\sum_{x:\,p(x,y)>0} x\,p_{X|Y}(x|y) & \text{in discrete case} \\[2ex] \displaystyle\int_{-\infty}^{\infty} x\,f_{X|Y}(x|y)\,dx & \text{in continuous case} \end{cases}$$

Observe that in either case, $E[X\,|Y=y]$ satisfies all properties of an expectation, e.g.

$$E[X_1+\ldots+X_n\,|Y=y] = E[X_1\,|Y=y]+\ldots+E[X_n\,|Y=y]$$

Thus, random variables $E[X_i\,|Y]$ satisfy

$$E[X_1+\ldots+X_n\,|Y] = E[X_1\,|Y]+\ldots+E[X_n\,|Y]$$

Also, for a function g,

$$E[g(Y)X\,|Y] = g(Y)E[X|Y]$$

as $E[g(y)X\,|Y=y] = g(y)E[X|Y=y]$. Finally, $E[g(Y)\,|\,Y] = g(Y)$

Conditional expectation: random variable $E[X|Y]$

$$E[X\,|Y=y] = \begin{cases} \displaystyle\sum_{x:\,p(x,y)>0} x\,p_{X|Y}(x|y) & \text{in discrete case} \\ \displaystyle\int_{-\infty}^{\infty} x\,f_{X|Y}(x|y)\,dx & \text{in continuous case} \end{cases}$$

Observe that in either case, $\varphi(y) = E[X\,|Y=y]$ is a function of y. Random variable $E[X|Y]$ is defined by letting $E[X|Y] = \varphi(Y)$.

Theorem.

$$E\Big[E[X|Y]\Big] = E[X]$$

Proof. Assume X and Y are discrete random variables.

$$E\Big[E[X|Y]\Big] = E\Big[\varphi(Y)\Big] = \sum_{y:\,p_{\mathrm{y}}(y)>0} \varphi(y)p_{\mathrm{y}}(y) = \sum_{y:\,p_{\mathrm{y}}(y)>0}\left(\sum_{x:\,p(x,y)>0} x\,p_{X|Y}(x|y)\right)p_{\mathrm{y}}(y)$$

$$= \sum_{y:\,p_{\mathrm{y}}(y)>0}\left(\sum_{x:\,p(x,y)>0} x\,\frac{p(x,y)}{p_{\mathrm{y}}(y)}\right)p_{\mathrm{y}}(y) = \sum_{y:\,p_{\mathrm{y}}(y)>0}\left(\sum_{x:\,p(x,y)>0} x\,p(x,y)\right) = \sum_{x,y:\,p(x,y)>0} x\,p(x,y) = E[X].$$

Conditional expectation: random variable $E[X|Y]$

$$E[X\,|Y=y] = \begin{cases} \displaystyle\sum_{x:\,p(x,y)>0} x\,p_{X|Y}(x|y) & \text{in discrete case} \\ \displaystyle\int_{-\infty}^{\infty} x\,f_{X|Y}(x|y)\,dx & \text{in continuous case} \end{cases}$$

Observe that in either case, $\varphi(y) = E[X\,|Y=y]$ is a function of y.
Random variable $E[X|Y]$ is defined by letting $E[X|Y] = \varphi(Y)$.

Theorem.

$$E\Big[E[X|Y]\Big] = E[X]$$

Proof. Assume X and Y are continuous random variables.

$$E\Big[E[X|Y]\Big] = E\Big[\varphi(Y)\Big] = \int_{-\infty}^{\infty} \varphi(y) f_{\mathsf{y}}(y)\,dy = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} x\,f_{X|Y}(x|y)\,dx\right) f_{\mathsf{y}}(y)\,dy$$

$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} x\,\frac{f(x,y)}{f_{\mathsf{y}}(y)}\,dx\right) f_{\mathsf{y}}(y)\,dy = \iint_{\mathbb{R}^2} x\,f(x,y)\,dx\,dy = E[X].$$

Conditional expectation: random variable $E[X|Y]$

Observe that in either case, $\varphi(y) = E[X\,|Y = y]$ is a function of y. Random variable $E[X|Y]$ is defined by letting $E[X|Y] = \varphi(Y)$.

Theorem.

$$E\Big[E[X|Y]\Big] = E[X]$$

Note that $E\Big[E[X|Y]\Big] = E[X]$ holds even when one of the variables is continuous and the other is discrete.

Example. Weather prediction. Let

$X =$ weather after tomorrow and $Y =$ weather tomorrow

Then,

$E[X] =$ today's prediction of weather after tomorrow

$E[X|Y] =$ tomorrow's prediction of weather after tomorrow

and

$E\Big[E[X|Y]\Big] =$ today's prediction of tomorrow's prediction of weather after tomorrow

Wald's equation.

Wald's equation. Suppose $X_1, X_2, \ldots$ are a sequence of random variables with finite mean $E[X_i] = \mu$ $(i \geq 1)$, and let N be a nonnegative integer-valued discrete random variable with finite mean $E[N]$. Assume N is independent from $X_1, X_2, \ldots$. Then,

$$E\left[\sum_{i=1}^{N} X_i\right] = \mu\, E[N], \qquad \text{where, in this notations } \sum_{i=1}^{0} X_i = 0.$$

Proof. For a given integer $n \geq 0$,

$$E\left[\sum_{i=1}^{N} X_i \,\middle|\, N = n\right] = E\left[\sum_{i=1}^{n} X_i \,\middle|\, N = n\right] = E\left[\sum_{i=1}^{n} X_i\right] = n\mu.$$

Therefore, $$E\left[\sum_{i=1}^{N} X_i \,\middle|\, N\right] = N\mu,$$ and

$$E\left[\sum_{i=1}^{N} X_i\right] = E\left[E\left[\sum_{i=1}^{N} X_i \,\middle|\, N\right]\right] = E[N\mu] = \mu\, E[N].$$

Wald's equation.

Wald's equation. Suppose $X_1, X_2, \ldots$ are a sequence of random variables with finite mean $E[X_i] = \mu$ $(i \geq 1)$, and let N be a nonnegative integer-valued discrete random variable with finite mean $E[N]$. Assume N is independent from $X_1, X_2, \ldots$. Then,

$$E\left[\sum_{i=1}^{N} X_i\right] = \mu\, E[N], \qquad \text{where, in this notations } \sum_{i=1}^{0} X_i = 0.$$

Alternative proof using indicator variables.

Observe that $\sum_{i=1}^{N} X_i = \sum_{i=1}^{\infty} X_i\, I_{N \geq i}$. Therefore,

$$E\left[\sum_{i=1}^{N} X_i\right] = E\left[\sum_{i=1}^{\infty} X_i\, I_{N\geq i}\right] = \sum_{i=1}^{\infty} E\left[X_i\, I_{N\geq i}\right] = \sum_{i=1}^{\infty} E[X_i] E\left[I_{N\geq i}\right]$$

$$= \sum_{i=1}^{\infty} \mu P(N \geq i) = \mu \sum_{i=1}^{\infty} P(N \geq i) = \mu\, E[N].$$

Wald's equation.

Wald's equation. Suppose $X_1, X_2, \dots$ are a sequence of random variables with finite mean $E[X_i] = \mu$ $(i \geq 1)$, and let N be a nonnegative integer-valued discrete random variable with finite mean $E[N]$. Assume N is independent from $X_1, X_2, \dots$. Then,

$$E\left[\sum_{i=1}^{N} X_i\right] = \mu\, E[N], \qquad \text{where, in this notations } \sum_{i=1}^{0} X_i = 0.$$

Example. Let $N =$ the number of customers per day

and $X_i =$ income from the i^{th} customer

Then, the total expected income per day equals

$$E\left[\sum_{i=1}^{N} X_i\right] = \mu\, E[N].$$

Conditional variance: random variable $Var(X|Y)$

Conditional variance is a random variable defined as

$$Var(X|Y) = E\Big[\big(X - E[X|Y]\big)^2 \,\Big|\, Y\Big]$$

Lemma. $Var(X|Y) = E[X^2|Y] - \big(E[X|Y]\big)^2$

Theorem (the law of total variance).

$$Var(X) = E\Big[Var(X|Y)\Big] + Var\Big(E[X|Y]\Big)$$

Proof.

$$E\Big[Var(X|Y)\Big] = E\Big[E[X^2|Y]\Big] - E\Big[\big(E[X|Y]\big)^2\Big] = E[X^2] - E\Big[\big(E[X|Y]\big)^2\Big]$$

$$Var\Big(E[X|Y]\Big) = E\Big[\big(E[X|Y]\big)^2\Big] - \Big(E\Big[E[X|Y]\Big]\Big)^2 = E\Big[\big(E[X|Y]\big)^2\Big] - \big(E[X]\big)^2$$

Therefore,

$$E\Big[Var(X|Y)\Big] + Var\Big(E[X|Y]\Big) = E[X^2] - \big(E[X]\big)^2 = Var(X)$$

Conditional variance. The law of total variance.

Theorem (the law of total variance).

$$Var(X) = E\Big[Var(X|Y)\Big] + Var\Big(E[X|Y]\Big)$$

Variance of a random sum of random variables.

Suppose $X_1, X_2, \ldots$ are independent random variables with finite

$$E[X_i] = \mu \quad \text{and} \quad Var(X_i) = \sigma^2,$$

and let N be a nonnegative integer-valued discrete random variable with finite mean and variance. Assume N is independent from $X_1, X_2, \ldots$. Then,

$$Var\left(\sum_{i=1}^{N} X_i\right) = \sigma^2 E[N] + \mu^2 Var(N).$$

Recall Wald's equation: $E\left[\sum_{i=1}^{N} X_i\right] = \mu\, E[N]$.

Variance of a random sum of random variables.

$$Var\left(\sum_{i=1}^{N} X_i\right) = \sigma^2\, E[N] + \mu^2\, Var(N).$$

Proof. Recall $E\left[\sum_{i=1}^{N} X_i \,\middle|\, N\right] = N\mu$. Thus,

$$Var\left(\sum_{i=1}^{N} X_i \,\middle|\, N\right) = E\left[\left(\sum_{i=1}^{N} X_i - N\mu\right)^2 \,\middle|\, N\right] = N\sigma^2 \quad \text{as}$$

$$E\left[\left(\sum_{i=1}^{N} X_i - N\mu\right)^2 \,\middle|\, N = n\right] = E\left[\left(\sum_{i=1}^{n} X_i - n\mu\right)^2\right] = \sum_{i=1}^{n} Var(X_i) = n\sigma^2$$

Therefore, $$Var\left(\sum_{i=1}^{N} X_i\right) = E\left[Var\left(\sum_{i=1}^{N} X_i \,\middle|\, N\right)\right] + Var\left(E\left[\sum_{i=1}^{N} X_i \,\middle|\, N\right]\right)$$

$$= E\left[N\sigma^2\right] + Var\left(N\mu\right) = \sigma^2\, E[N] + \mu^2\, Var(N).$$

Conditional expectation as a projection.

Theorem (the law of total variance).

$$Var(X) = E\Big[Var(X|Y)\Big] + Var\Big(E[X|Y]\Big)$$

Consider the following distance: for a pair of random variables, U and V, let

$$d(U,V) = \sqrt{E\Big[(U-V)^2\Big]}.$$

Pythagorean Theorem. Random variable $\varphi(Y) = E[X|Y]$ is an orthogonal projection onto a subspace of random variables

$$\mathcal{F}(Y) = \Big\{g(Y) \,:\, \forall\, g : \mathbb{R} \to \mathbb{R}\Big\}.$$

That is, for any function g,

$$d^2\Big(X, g(Y)\Big) = d^2\Big(X, \varphi(Y)\Big) + d^2\Big(\varphi(Y), g(Y)\Big)$$

Conditional variance. The law of total variance.

Pythagorean Theorem. Random variable $\varphi(Y) = E[X|Y]$ is an orthogonal projection onto a subspace of random variables

$$\mathcal{F}(Y) = \Big\{ g(Y) \, : \, \forall \, g : \mathbb{R} \to \mathbb{R} \Big\}.$$

That is, for any function g,

$$d^2\Big(X, g(Y)\Big) = d^2\Big(X, \varphi(Y)\Big) + d^2\Big(\varphi(Y), g(Y)\Big)$$

Proof. Let $Z = X - g(Y)$. Recall that

$$E\Big[Var(Z|Y)\Big] = E\Big[E[Z^2|Y]\Big] - E\Big[\Big(E[Z|Y]\Big)^2\Big] = E[Z^2] - E\Big[\Big(E[Z|Y]\Big)^2\Big]$$

$$= E\Big[\Big(X - g(Y)\Big)^2\Big] - E\Big[\Big(\varphi(Y) - g(Y)\Big)^2\Big] = d^2\Big(X, g(Y)\Big) - d^2\Big(\varphi(Y), g(Y)\Big)$$

as $E[Z|Y] = E[X|Y] - g(Y) = \varphi(Y) - g(Y)$. On the other hand,

$$E\Big[Var(Z|Y)\Big] = E\Big[E\Big[(Z - E[Z|Y] \,\Big|\, Y\Big]\Big] = E\Big[E\Big[(X - g(Y) - E[Z|Y] \,\Big|\, Y\Big]\Big]$$

$$= E\Big[\Big(X - \varphi(Y)\Big)^2\Big] = d^2\Big(X, \varphi(Y)\Big) \qquad \text{as} \quad E[g(Y)|Y] = g(Y).$$

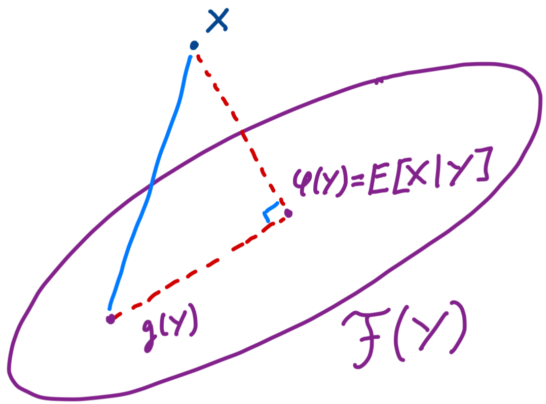
Conditional expectation as a projection.

Pythagorean Theorem. Random variable $\varphi(Y) = E[X|Y]$ is an orthogonal projection onto a subspace of random variables

$$\mathcal{F}(Y) = \Big\{ g(Y) \,:\, \forall\, g : \mathbb{R} \to \mathbb{R} \Big\}.$$

That is, for any function g,

$$d^2\Big(X, g(Y)\Big) = d^2\Big(X, \varphi(Y)\Big) + d^2\Big(\varphi(Y), g(Y)\Big)$$



Thus, $d^2\Big(X, g(Y)\Big) \geq d^2\Big(X, \varphi(Y)\Big) \quad \forall g : \mathbb{R} \to \mathbb{R}$. In other words, $\varphi(Y) = E[X|Y]$ is an orthogonal projection onto $\mathcal{F}(Y)$.

Conditional expectation as a projection.

Pythagorean Theorem. Random variable $\varphi(Y) = E[X|Y]$ is an orthogonal projection onto a subspace of random variables

$$\mathcal{F}(Y) = \Big\{ g(Y) \,:\, \forall\, g : \mathbb{R} \to \mathbb{R} \Big\}.$$

That is, for any function g,

$$d^2\Big(X, g(Y)\Big) = d^2\Big(X, \varphi(Y)\Big) + d^2\Big(\varphi(Y), g(Y)\Big)$$

Consider linear subspaces $R \subseteq F$, and the respective orthogonal projections, Π_R and Π_F. Then,

$$\Pi_R \Pi_F = \Pi_R.$$

The space of all real-valued constant functions $\mathbb{R} \subseteq \mathcal{F}(Y)$. We have shown that $E[X|Y] = \Pi_F X$ is an orthogonal projection. Similarly, $E[X] = \Pi_R X$. Thus,

$$\Pi_R \Pi_F = \Pi_R \quad \Leftrightarrow \quad E\Big[E[X|Y]\Big] = E[X].$$

Also,

$$\Pi_F Z = Z \quad \forall Z \in \mathcal{F}(Y) \quad \Leftrightarrow \quad E[g(Y)\,|\,Y] = g(Y).$$

The law of total variance via Pythagorean Theorem.

$$d(U,V) = \sqrt{E\big[(U-V)^2\big]}$$

Pythagorean Theorem.

$$d^2\big(X, g(Y)\big) = d^2\big(X, \varphi(Y)\big) + d^2\big(\varphi(Y), g(Y)\big)$$

The law of total variance, $Var(X) = E\big[Var(X|Y)\big] + Var\big(E[X|Y]\big)$ is equivalent to

$$d^2\big(X, 0\big) = d^2\big(X, \varphi(Y)\big) + d^2\big(\varphi(Y), 0\big)$$

Indeed,

$$d^2\big(X,0\big) = d^2\big(X,\varphi(Y)\big) + d^2\big(\varphi(Y),0\big) \quad \Leftrightarrow \quad E[X^2] = E\Big[\big(X-\varphi(Y)\big)^2\Big] + E\big[\varphi^2(Y)\big]$$

$$\Leftrightarrow \quad E[X^2] - \big(E[X]\big)^2 = E\Big[E\Big[\big(X-\varphi(Y)\big)^2 \,\Big|\, Y\Big]\Big] + E\big[\varphi^2(Y)\big] - \Big(E\big[E[X|Y]\big]\Big)^2$$

$$\Leftrightarrow \quad Var(X) = E\big[Var(X|Y)\big] + Var\big(E[X|Y]\big)$$

Conditional distributions: randomization formula.

Definition. Suppose X and Y are discrete random variables. For a given y such that $p_{\mathsf{y}}(y) > 0$, the conditional probability mass function $p_{X|Y}(x|y)$ is defined as

$$p_{X|Y}(x|y) = P\Big(X = x \,\Big|\, Y = y\Big) = \frac{p(x,y)}{p_{\mathsf{y}}(y)} \qquad \forall x \text{ s.t. } p(x,y) > 0.$$

We have the following randomization formula

$$p_{\mathsf{x}}(x) = \sum_{y:\, p(x,y)>0} p(x,y) = \sum_{y:\, p(x,y)>0}^{\infty} p_{X|Y}(x|y) p_{\mathsf{y}}(y).$$

Definition. Suppose X and Y are continuous random variables. For a given y such that $f_{\mathsf{y}}(y) > 0$, the conditional probability density function $f_{X|Y}(x|y)$ is defined as

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_{\mathsf{y}}(y)} \qquad \forall x \in \mathbb{R}.$$

The following randomization formula holds

$$f_{\mathsf{x}}(x) = \int_{-\infty}^{\infty} f(x,y)\, dy = \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_{\mathsf{y}}(y)\, dy.$$

Conditional distributions: randomization formula.

Example. Suppose $X_1, X_2, \ldots$ are i.i.d. Bernoulli random variables with parameter p, and N be a Poisson random variable with parameter $\lambda > 0$. Assume $N, X_1, X_2, \ldots$ are independent. Let

$$Y = \sum_{i=1}^{N} X_i, \qquad \text{where, in this notations } \sum_{i=1}^{0} X_i = 0.$$

Find the probability mass function p_{y}.

Solution: Fix an integer $n = 0, 1, \ldots$. Then, the conditional probability mass function

$$p_{Y|N}(k|n) = P\left(\sum_{i=1}^{N} X_i = k \,\middle|\, N = n\right) = P\left(\sum_{i=1}^{n} X_i = k\right) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, \ldots, n$, as $\sum_{i=1}^{n} X_i$ is binomial with parameters (n, p).

Conditional distributions: randomization formula.

Solution (continued): Fix an integer $n = 0, 1, \ldots$. Then,

$$p_{Y|N}(k|n) = P\left(\sum_{i=1}^{N} X_i = k \,\middle|\, N = n\right) = P\left(\sum_{i=1}^{n} X_i = k\right) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, \ldots, n$, as $\sum_{i=1}^{n} X_i$ is binomial with parameters (n, p).

Recall the randomization formula:

$$p_{\mathsf{X}}(x) = \sum_{y:\, p(x,y)>0} p_{X|Y}(x|y) p_{\mathsf{Y}}(y).$$

Thus, for $k = 0, 1, 2, \ldots$, we have a randomization formula

$$p_{\mathsf{Y}}(k) = \sum_{n=k}^{\infty} p_{Y|N}(k|n) p_N(n) = \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} e^{-\lambda} \frac{\lambda^n}{n!}$$

$$= e^{-\lambda} \frac{(\lambda p)^k}{k!} \sum_{n=k}^{\infty} \frac{\big(\lambda(1-p)\big)^{n-k}}{(n-k)!} = e^{-\lambda} \frac{(\lambda p)^k}{k!} e^{\lambda(1-p)} = e^{-\lambda p} \frac{(\lambda p)^k}{k!}.$$

Conditional distributions.

Definition. Suppose X is a discrete random variable with probability mass function p_{x} and Y is a continuous random variable with probability density function f_{y}. For a given y such that $f_{\mathsf{y}}(y) > 0$, the conditional probability mass function $p_{X|Y}(x|y)$ is defined as

$$p_{X|Y}(x|y) = \frac{\frac{d}{dy}P\Big(X = x \cap Y \le y\Big)}{f_{\mathsf{y}}(y)} \qquad \forall x \text{ s.t. } p_{\mathsf{x}}(x) > 0.$$

Definition. Suppose X is a continuous random variable with probability density function f_{x} and Y is a discrete random variable with probability mass function p_{y}. For a given y such that $p_{\mathsf{y}}(y) > 0$, the conditional probability density function $f_{X|Y}(x|y)$ is defined as

$$f_{X|Y}(x|y) = \frac{\frac{d}{dx}P\Big(X \le x \cap Y = y\Big)}{p_{\mathsf{y}}(y)} = \frac{d}{dx}P\Big(X \le x \,\Big|\, Y = y\Big) \quad \forall x \in \mathbb{R}.$$

Property: $f_{X|Y}(x|y)p_{\mathsf{y}}(y) = \frac{d}{dx}P\Big(X \le x \cap Y = y\Big) = p_{Y|X}(y|x)f_{\mathsf{x}}(x)$.

Conditional distributions.

Definition. Suppose X is a discrete random variable with probability mass function p_{x} and Y is a continuous random variable with probability density function f_{y}. For a given y such that $f_{\mathsf{y}}(y) > 0$, the conditional probability mass function $p_{X|Y}(x|y)$ is defined as

$$p_{X|Y}(x|y) = \frac{\frac{d}{dy} P\Big(X = x \cap Y \leq y\Big)}{f_{\mathsf{y}}(y)} \qquad \forall x \text{ s.t. } p_{\mathsf{x}}(x) > 0.$$

The probability mass function is computed as follows:

$$p_{\mathsf{x}}(x) = \int_{-\infty}^{\infty} p_{X|Y}(x|y)\, f_{\mathsf{y}}(y)\, dy$$

Example. Suppose X is a discrete random variable and Y is an exponential random variable with parameter $\lambda > 0$. For a given $y > 0$, the conditional probability mass function $p_{X|Y}(x|y)$ is Poisson with parameter y:

$$p_{X|Y}(k|y) = e^{-y} \frac{y^k}{k!} \qquad k = 0, 1, 2, \ldots$$

Find the probability mass function $p_{\mathsf{x}}(k)$.

Conditional distributions.

Example. Suppose X is a discrete random variable and Y is an exponential random variable with parameter $\lambda > 0$. For a given $y > 0$, the conditional probability mass function $p_{X|Y}(x|y)$ is Poisson with parameter y:

$$p_{X|Y}(k|y) = e^{-y}\frac{y^k}{k!} \qquad k = 0, 1, 2, \dots$$

Find the probability mass function $p_{\mathsf{x}}(k)$.

Solution: For $k = 0, 1, 2, \dots$,

$$p_{\mathsf{x}}(k) = \int_0^\infty e^{-y}\frac{y^k}{k!}\,\lambda e^{-\lambda y}\,dy = \frac{\lambda}{(\lambda+1)^{k+1}k!}\int_0^\infty \big((\lambda+1)y\big)^k e^{-(\lambda+1)y}\,(\lambda+1)dy$$

$$= \frac{\lambda}{(\lambda+1)^{k+1}k!}\int_0^\infty z^k e^{-z}\,dz = \frac{\lambda}{(\lambda+1)^{k+1}} = \frac{\lambda}{\lambda+1}\left(\frac{1}{\lambda+1}\right)^k$$

as we let $z = (\lambda+1)y$.

Thus, $X+1$ is a geometric random variable with parameter $p = \frac{\lambda}{\lambda+1}$.

Conditional distributions.

Definition. Suppose X is a continuous random variable with probability density function f_{x} and Y is a discrete random variable with probability mass function p_{y}. For a given y such that $p_{\mathsf{y}}(y) > 0$, the conditional probability density function $f_{X|Y}(x|y)$ is defined as

$$f_{X|Y}(x|y) = \frac{\frac{d}{dx}P\Big(X \le x \cap Y = y\Big)}{p_{\mathsf{y}}(y)} = \frac{d}{dx}P\Big(X \le x \,\Big|\, Y = y\Big) \quad \forall x \in \mathbb{R}.$$

The probability density function is computed as follows:

$$f_{\mathsf{x}}(x) = \sum_{y:\, p_{\mathsf{y}}(y)>0} f_{X|Y}(x|y)\, p_{\mathsf{y}}(y)$$

Example. Suppose X is a continuous random variable and Y is geometric random variable with parameter $p \in (0,1)$. For a given $m = 1, 2, \ldots$, the conditional probability density function $f_{X|Y}(x|m)$ is Gamma with parameters (m, λ):

$$f_{X|Y}(x|m) = \frac{1}{\Gamma(m)}\lambda^m x^{m-1} e^{-\lambda x} \quad \forall x > 0, \quad \text{where} \quad \Gamma(m) = (m-1)!$$

Find the probability density function $f_{\mathsf{x}}(x)$.

Conditional distributions.

Example. Suppose X is a continuous random variable and Y is geometric random variable with parameter $p \in (0,1)$. For a given $m = 1, 2, \ldots$, the conditional probability density function $f_{X|Y}(x|m)$ is Gamma with parameters (m, λ):

$$f_{X|Y}(x|m) = \frac{1}{\Gamma(m)} \lambda^m x^{m-1} e^{-\lambda x} \quad \forall x > 0, \quad \text{where} \quad \Gamma(m) = (m-1)!$$

Find the probability density function $f_{\mathsf{x}}(x)$.

Solution: For $x > 0$,

$$f_{\mathsf{x}}(x) = \sum_{m=1}^{\infty} f_{X|Y}(x|m)\, p_{\mathsf{y}}(m) = \sum_{m=1}^{\infty} \frac{1}{(m-1)!} \lambda^m x^{m-1} e^{-\lambda x}\, p\,(1-p)^{m-1}$$

$$= \lambda p e^{-\lambda x} \sum_{m=1}^{\infty} \frac{1}{(m-1)!} \big(\lambda(1-p)x\big)^{m-1} = \lambda p e^{-\lambda x} e^{\lambda(1-p)x} = \lambda p e^{-\lambda p x}.$$

Thus, X is an exponential random variable with parameter λp.

Conditional distributions.

Example. Suppose X is a continuous random variable and Y is geometric random variable with parameter $p \in (0,1)$. For a given $m = 1,2,\ldots$, the conditional probability density function $f_{X|Y}(x|m)$ is Gamma with parameters (m,λ):

$$f_{X|Y}(x|m) = \frac{1}{\Gamma(m)}\lambda^m x^{m-1} e^{-\lambda x} \quad \forall x > 0, \quad \text{where} \quad \Gamma(m) = (m-1)!$$

Find the probability density function $f_{\mathsf{x}}(x)$.

Answer: For $x > 0$,

$$f_{\mathsf{x}}(x) = \lambda p e^{-\lambda p x}.$$

Thus, X is an exponential random variable with parameter λp.

This example may come up when $X_1, X_2, \ldots$ are i.i.d. exponential random variables with parameter $\lambda > 0$, random variable Y (geometric with parameter p) is independent of $X_1, X_2, \ldots$, and

$$X = \sum_{i=1}^{Y} X_i, \quad \text{where, in this notations} \sum_{i=1}^{0} X_i = 0.$$

Joint cumulative distribution function.

Definition. For a pair of random variables X and Y, the function

$$F_{\mathrm{x,y}}(x,y) = P(X \leq x \cap Y \leq y), \qquad x, y \in \mathbb{R},$$

is the joint cumulative distribution function.

Example. Suppose X and Y are continuous random variables, then

$$F_{\mathrm{x,y}}(a,b) = P(X \leq a \cap Y \leq b) = \int_{-\infty}^{b} \int_{-\infty}^{a} f(x,y)\, dx\, dy$$

and

$$f(a,b) = \frac{\partial^2}{\partial a\, \partial b} F_{\mathrm{x,y}}(a,b) = \frac{\partial^2}{\partial b\, \partial a} F_{\mathrm{x,y}}(a,b).$$

Similarly, for random variables $X_1, X_2, \ldots, X_n$, the function

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1 \cap X_2 \leq x_2 \cap \ldots \cap X_n \leq x_n), \quad x_1, x_2, \ldots, x_n \in \mathbb{R},$$

is the joint cumulative distribution function.

Moment generating functions.

Definition. For a given random variable X, the function

$$M_X(s) = E\left[e^{sX}\right]$$

is called the moment generating function (m.g.f.).

Properties: • $M_X(0) = 1$.

• $$M_X(s) = E\left[e^{sX}\right] = \begin{cases} \sum\limits_{x:\, p_X(x)>0} e^{sx} p_X(x) & \text{if } X \text{ is discrete,} \\ \int\limits_{-\infty}^{\infty} e^{sx} f_X(x)\, dx & \text{if } X \text{ is continuous.} \end{cases}$$

• The derivatives of $M_X(s)$ are computed as follows

$$M_X'(s) = \frac{d}{ds} E[e^{sX}] = E[Xe^{sX}] \quad \text{and}$$

$$M_X^{(n)}(s) = \frac{d^n}{ds^n} E[e^{sX}] = E\left[\frac{d^n}{ds^n} e^{sX}\right] = E[X^n e^{sX}].$$

Thus, $M_X^{(n)}(0) = E[X^n]$ (the n^{th} moment), and

$$E[X] = M_X'(0), \quad E[X^2] = M_X''(0), \quad Var(X) = M_X''(0) - \big(M_X'(0)\big)^2.$$

Moment generating functions.

Definition. For a given random variable X, the function

$$M_X(s) = E\left[e^{sX}\right]$$

is called the moment generating function (m.g.f.).

An important property of $M_X(s)$: If X and Y are independent random variables with the respective moment generating functions $M_X(s)$ and $M_Y(s)$, then the moment generating function of $X + Y$ is

$$M_{X+Y}(s) = E\left[e^{s(X+Y)}\right] = E\left[e^{sX}e^{sY}\right] = E\left[e^{sX}\right]E\left[e^{sY}\right] = M_X(s)\, M_Y(s).$$

Hence, if $X_1, X_2, \ldots, X_n$ are independent random variables, then the moment generating function of $X = X_1 + X_2 + \ldots + X_n$ equals

$$M_X(s) = M_{X_1}(s) \cdot M_{X_2}(s) \cdot \ldots \cdot M_{X_n}(s).$$

Moment generating functions.

Example. Consider a Bernoulli random variable X with parameter $p \in [0, 1]$, i.e., $X \sim Bernoulli(p)$. Then,

$$M_X(s) = E[e^{sX}] = \sum_{k=0,1} e^{sk} p_{\mathsf{X}}(k) = 1 \cdot (1-p) + e^s \cdot p.$$

Hence,

$$M_X(s) = 1 - p + pe^s \quad \text{with the domain} \quad s \in \mathbb{R}.$$

Example. Consider a binomial random variable X with parameters (n, p), i.e., $X \sim Binomial(n, p)$. Then,

$$X = X_1 + X_2 + \ldots + X_n,$$

where $X_1, X_2, \ldots, X_n$ are independent $Bernoulli(p)$ random variables. Thus,

$$M_X(s) = M_{X_1}(s) \cdot M_{X_2}(s) \cdot \ldots \cdot M_{X_n}(s) = \Big(1 - p + pe^s\Big)^n, \qquad s \in \mathbb{R}.$$

Hence, $E[X] = M'_X(0) = np, \quad E[X^2] = M''_X(0) = np + n(n-1)p^2,$

$$\text{and} \qquad Var(X) = E[X^2] - \big(E[X]\big)^2 = np(1-p).$$

Moment generating functions.

Example. Consider a binomial random variable X with parameters (n, p), i.e., $X \sim Binomial(n, p)$. Then,

$$M_X(s) = \left(1 - p + pe^s\right)^n, \qquad s \in \mathbb{R}.$$

Alternative derivation via Binomial Theorem:

$$M_X(s) = \sum_{k=0}^{n} e^{sk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} \left(pe^s\right)^k (1-p)^{n-k} = \left(1 - p + pe^s\right)^n$$

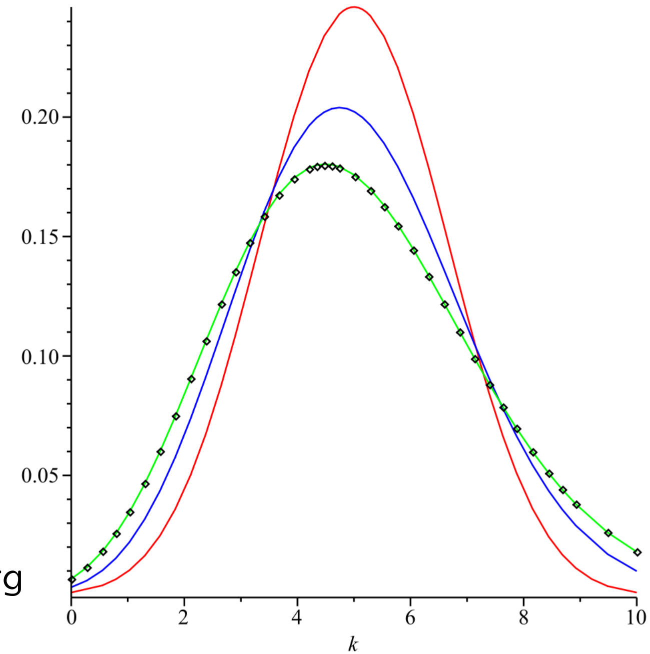
Example. Consider a Poisson random variable X with parameter $\lambda > 0$. i.e., $X \sim Poisson(\lambda)$. Then,

$$M_X(s) = E[e^{sX}] = \sum_{k=0}^{\infty} e^{sk} p_{\mathsf{X}}(k) = \sum_{k=0}^{\infty} e^{sk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\left(\lambda e^s\right)^k}{k!} = e^{-\lambda} e^{\lambda e^s}.$$

Hence,

$$M_X(s) = \exp\left\{\lambda(e^s - 1)\right\}, \qquad s \in \mathbb{R}.$$

Poisson vs Binomial.



Picture credit: Wikipedia.org

Dots: Poisson($\lambda = 5$) Red: Binomial($n = 10,\ p = \frac{1}{2}$)

Blue: Binomial($n = 20,\ p = \frac{1}{4}$)

Green: Binomial($n = 1000,\ p = \frac{1}{200}$)

Poisson vs Binomial.

Let $\lambda > 0$ be given. Suppose Y is a Poisson random variable with parameter λ and S_n is a Binomial random variable with parameters n and $p = \frac{\lambda}{n}$.

• **Theorem.** For a given integer $k \geq 0$, $\lim_{n\to\infty} P(S_n = k) = P(Y = k)$.

Thus, for n large enough, $P(S_n = k) \approx P(Y = k)$.

Alternative proof: $\forall s \in \mathbb{R}$,

$$M_{S_n}(s) = \left(1 - p + pe^s\right)^n = \left(1 - \frac{\lambda}{n} + \frac{\lambda}{n}e^s\right)^n = \left(1 + \frac{\lambda(e^s - 1)}{n}\right)^n$$

Hence,

$$\lim_{n\to\infty} M_{S_n}(s) = \lim_{n\to\infty} \left(1 + \frac{\lambda(e^s - 1)}{n}\right)^n = e^{\lambda(e^s-1)} = M_Y(s).$$

Theorem. The cumulative distribution function $F_X(x)$ is unique for a m.g.f. $M_X(s)$. Moreover, if $\lim_{n\to\infty} M_{X_n}(s) = M_X(s)$, then the cumulative distribution functions also converge, i.e.,

$$\lim_{n\to\infty} F_{X_n}(a) = F_X(a) \qquad \forall a \in \mathbb{R}$$

Moment generating functions.

Example. Consider a geometric random variable X with parameter $p \in (0,1)$, i.e., $X \sim Geometric(p)$. Then,

$$M_X(s) = E[e^{sX}] = \sum_{k=1}^{\infty} e^{sk} p_{\mathsf{x}}(k) = \sum_{k=1}^{\infty} e^{sk}(1-p)^{k-1} p$$

$$= pe^s \sum_{k=1}^{\infty} \Big((1-p)e^s\Big)^{k-1} = \frac{pe^s}{1-(1-p)e^s} \quad \text{when} \quad (1-p)e^s < 1.$$

Hence,

$$M_X(s) = \frac{pe^s}{1-(1-p)e^s}, \quad s \in \Big(-\infty, -\ln(1-p)\Big).$$

Differentiating $M_X(s) = \frac{pe^s}{1-(1-p)e^s}$ we obtain

$$M_X'(s) = \frac{pe^s}{\Big(1-(1-p)e^s\Big)^2}, \quad M_X''(s) = \frac{pe^s + p(1-p)e^{2s}}{\Big(1-(1-p)e^s\Big)^3}.$$

Therefore, $E[X] = M_X'(0) = \frac{1}{p}$, $E[X^2] = M_X''(0) = \frac{2-p}{p^2}$, and

$$Var(X) = E[X^2] - \Big(E[X]\Big)^2 = \frac{1-p}{p^2}.$$

Moment generating function for $X \sim Exponential(\lambda)$

Example. Consider a exponential random variable X with parameter $\lambda > 0$, i.e., $X \sim Exponential(\lambda)$. Then, for $s < \lambda$,

$$M_X(s) = \int_0^\infty e^{sx} \lambda e^{-\lambda x}\, dx = \frac{\lambda}{\lambda - s} \int_0^\infty (\lambda - s) e^{-(\lambda - s)x}\, dx$$

Let $y = (\lambda - s)x$, then

$$M_X(s) = \frac{\lambda}{\lambda - s} \int_0^\infty e^{-y}\, dy = \frac{\lambda}{\lambda - s}, \qquad s \in (-\infty, \lambda).$$

Here,

$$M_X^{(n)}(s) = \frac{n!\lambda}{(\lambda - s)^{n+1}} \quad \text{implies} \quad E[X^n] = M_X^{(n)}(0) = \frac{n!}{\lambda^n},$$

and therefore, $E[X] = \frac{1}{\lambda}$ and $Var(X) = \frac{1}{\lambda^2}$.

Moment generating function for $X \sim Gamma(\alpha, \lambda)$

Example. Consider a gamma random variable X with parameters (α, λ), i.e., $X \sim Gamma(\alpha, \lambda)$. Then, for $s < \lambda$,

$$M_X(s) = \frac{1}{\Gamma(\alpha)} \int_0^\infty e^{sx} \lambda \big(\lambda x\big)^{\alpha-1} e^{-\lambda x} dx = \left(\frac{\lambda}{\lambda - s}\right)^\alpha \frac{1}{\Gamma(\alpha)} \int_0^\infty (\lambda - s)\big((\lambda - s)x\big)^{\alpha-1} e^{-(\lambda-s)x}\, dx$$

Let $y = (\lambda - s)x$, then

$$M_X(s) = \left(\frac{\lambda}{\lambda - s}\right)^\alpha \frac{1}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y}\, dy = \left(\frac{\lambda}{\lambda - s}\right)^\alpha, \qquad s \in (-\infty, \lambda).$$

Here,

$$M_X^{(n)}(s) = \frac{\alpha(\alpha+1)\dots(\alpha+n-1)\lambda^\alpha}{(\lambda - s)^{\alpha+n}} = \frac{\Gamma(\alpha+n)\lambda^\alpha}{\Gamma(\alpha)(\lambda - s)^{\alpha+n}}.$$

Hence,

$$E[X^n] = M_X^{(n)}(0) = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)\lambda^n}.$$

Therefore, $E[X] = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\lambda} = \frac{\alpha}{\lambda}$ and $Var(X) = \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}$.

Moment generating function for $X \sim Gamma(\alpha, \lambda)$.

• If X and Y are independent gamma random variables with the respective parameters (α, λ) and (β, λ). Their sum $X+Y$ is a gamma random variable with parameters $(\alpha+\beta, \lambda)$.

Alternative derivation: the moment generating functions are

$$M_X(s) = \left(\frac{\lambda}{\lambda - s}\right)^{\alpha} \quad \text{and} \quad M_Y(s) = \left(\frac{\lambda}{\lambda - s}\right)^{\beta}, \qquad s < \lambda.$$

By independence of X and Y,

$$M_{X+Y}(s) = M_X(s)\, M_Y(s) = \left(\frac{\lambda}{\lambda - s}\right)^{\alpha+\beta}, \qquad s < \lambda.$$

Since the cumulative distribution function $F_{X+Y}(x)$ of $X+Y$ is uniquely determined by the m.g.f. $M_{X+Y}(s)$, the sum $X+Y$ is a gamma random variable with parameters $(\alpha+\beta, \lambda)$.

• Let $X_1, X_2, \ldots$ be independent exponential random variables with parameter $\lambda > 0$. Then $T_n = \sum_{k=1}^{n} X_k$ $(n = 1, 2, \ldots)$ is a gamma random variable with parameters (n, λ).

Alternative derivation:

$$M_{T_n}(s) = M_{X_1}(s) \cdot \ldots \cdot M_{X_n}(s) = \left(\frac{\lambda}{\lambda - s}\right)^{n}.$$

Moment generating functions.

Example. Consider a standard normal random variable Z, i.e., $Z \sim \mathcal{N}(0,1)$. Then, its moment generating function equals

$$M_Z(s) = E[e^{sZ}] = \int_{-\infty}^{\infty} e^{sx} f(x)\ dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{sx} e^{-\frac{1}{2}x^2}\ dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2-2sx)}\ dx = e^{\frac{1}{2}s^2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-s)^2}\ dx$$

Hence,

$$M_Z(s) = \exp\left\{\frac{s^2}{2}\right\}, \qquad s \in \mathbb{R}.$$

Theorem. The cumulative distribution function $F_X(x)$ is unique for a m.g.f. $M_X(s)$. Moreover, if $\lim_{n\to\infty} M_{X_n}(s) = M_X(s)$, then the cumulative distribution functions also converge, i.e.,

$$\lim_{n\to\infty} F_{X_n}(a) = F_X(a) \qquad \forall a \in \mathbb{R}$$

Central Limit Theorem.

• **Central Limit Theorem (CLT).** Let $X_1, X_2, \ldots$ be i.i.d. random variables with mean μ and variance σ^2. Then,

$$\lim_{n\to\infty} P(a \le Y_n \le b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx \quad \Leftrightarrow \quad \lim_{n\to\infty} F_{Y_n}(a) = \Phi(a),$$

where $Y_n = \frac{X_1+X_2+\ldots+X_n-n\mu}{\sqrt{n}\sigma}$ and $\Phi(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx$ is the standard normal cumulative distribution function.

The de Moivre-Laplace Theorem is a case of CLT when $X_1, X_2, \ldots$ are independent Bernoulli random variables with the same parameter $p \in (0,1)$.

• **de Moivre-Laplace Theorem.** Let S_n be a binomial random variable with parameters (n,p), then

$$\lim_{n\to\infty} F_{Y_n}(a) = \Phi(a), \qquad \text{where} \quad Y_n = \frac{S_n - np}{\sqrt{np(1-p)}}.$$

Thus, it is sufficient to show that

$$\lim_{n\to\infty} M_{Y_n}(s) = \exp\left\{\frac{s^2}{2}\right\} \qquad \text{- m.g.f. for } \mathcal{N}(0,1).$$

de Moiver-Laplace Theorem via m.g.f.

Proof. Consider $S_n \sim Binomial(n,p)$ and let $Y_n = \frac{S_n - np}{\sqrt{np(1-p)}}$.

Then, $E[Y_n] = 0$ and $Var(Y_n) = 1$.

The moment generating function

$$M_{Y_n}(s) = \exp\left\{-s\frac{np}{\sqrt{np(1-p)}}\right\} \cdot M_{S_n}\left(\frac{s}{\sqrt{np(1-p)}}\right)$$

$$= \exp\left\{-s\frac{np}{\sqrt{np(1-p)}}\right\} \cdot \left(1 - p\left[1 - \exp\left\{\frac{s}{\sqrt{np(1-p)}}\right\}\right]\right)^n$$

and

$$\ln M_{Y_n}(s) = -s\frac{np}{\sqrt{np(1-p)}} + n \ln\left(1 - p\left[1 - \exp\left\{\frac{s}{\sqrt{np(1-p)}}\right\}\right]\right)$$

de Moiver-Laplace Theorem via m.g.f.

Proof (cont.): $S_n \sim Binomial(n,p)$ and $Y_n = \frac{S_n - np}{\sqrt{np(1-p)}}$.

$$\ln M_{Y_n}(s) = -s\frac{np}{\sqrt{np(1-p)}} + n\ln\left(1 - p\left[1 - \exp\left\{\frac{s}{\sqrt{np(1-p)}}\right\}\right]\right).$$

Here,

$$\alpha = 1 - \exp\left\{\frac{s}{\sqrt{np(1-p)}}\right\} = -\frac{s}{\sqrt{np(1-p)}} - \frac{s^2}{2np(1-p)} + O\left(\frac{1}{n^{3/2}}\right)$$

and therefore,

$$\ln(1-p\alpha) = -p\alpha - \frac{p^2\alpha^2}{2} + O\left(\frac{1}{n^{3/2}}\right) = \frac{ps}{\sqrt{np(1-p)}} + \frac{s^2}{2n(1-p)} - \frac{ps^2}{2n(1-p)} + O\left(\frac{1}{n^{3/2}}\right)$$

$$= \frac{ps}{\sqrt{np(1-p)}} + \frac{s^2}{2n} + O\left(\frac{1}{n^{3/2}}\right)$$

de Moiver-Laplace Theorem via m.g.f.

Proof (cont.): $S_n \sim Binomial(n, p)$ and $Y_n = \frac{S_n - np}{\sqrt{np(1-p)}}$.

$$\ln M_{Y_n}(s) = -s\frac{np}{\sqrt{np(1-p)}} + n\ln(1-p\alpha),$$

where

$$\alpha = 1-\exp\left\{\frac{s}{\sqrt{np(1-p)}}\right\} = -\frac{s}{\sqrt{np(1-p)}} - \frac{s^2}{2np(1-p)} + O\left(\frac{1}{n^{3/2}}\right)$$

and

$$\ln(1-p\alpha) = \frac{ps}{\sqrt{np(1-p)}} + \frac{s^2}{2n} + O\left(\frac{1}{n^{3/2}}\right)$$

Thus,

$$\ln M_{Y_n}(s) = \frac{s^2}{2} + O\left(\frac{1}{n^{1/2}}\right) \longrightarrow \frac{s^2}{2} \quad \text{as} \quad n \to \infty.$$

and

$$\lim_{n\to\infty} M_{Y_n}(s) = \exp\left\{\frac{s^2}{2}\right\} \quad \text{- m.g.f. for } \mathcal{N}(0,1).$$

Hence, $\lim_{n\to\infty} F_{Y_n}(a) = \Phi(a)$.