MTH 464/564

Lectures 18-24

Yevgeniy Kovchegov
Oregon State University

- Moment generating functions.
- Proof of de Moiver-Laplace Theorem.
- Proof of Central Limit Theorem.
- Chernoff bound.
- Jensen's inequality.
- Characteristic and generating functions.

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: • $M_X(0) = 1$.

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

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

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

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

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

$$E[X] = M'_X(0), \quad E[X^2] = M''_X(0), \quad Var(X) = M''_X(0) - \left(M'_X(0)\right)^2.$$

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,

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

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:

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

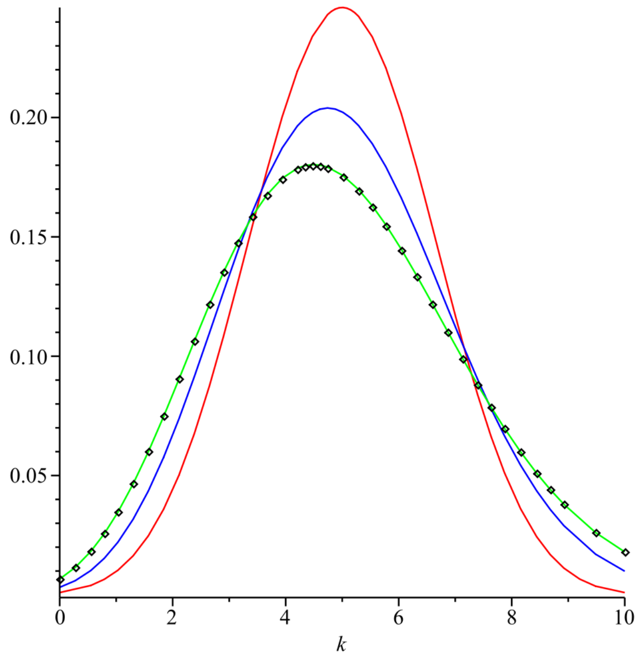
Example. Consider a Poisson random variable X with parameter $\lambda > 0$. i.e., $X \sim Poisson(\lambda)$. Then,

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

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,

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

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

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,

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

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 $X_1, X_2, \dots$ be independent exponential random variables with parameter $\lambda > 0$. Then $T_n = \sum_{k=1}^{n} X_k$ $(n = 1, 2, \dots)$ 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 $Y_n = \frac{S_n - np}{\sqrt{np(1-p)}}$.

$$\text{Then,} \quad E[Y_n] = 0 \quad \text{and} \quad Var(Y_n) = 1.$$

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

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,

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

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

Proof. It is sufficient to show that $\lim_{n\to\infty} M_{Y_n}(s) = \exp\left\{\frac{s^2}{2}\right\}$.

Let $M(s)$ denote the moment generating function of $\frac{X_j-\mu}{\sigma}$. Then,
$M(0) = 1$, $M'(0) = E\left[\frac{X_j-\mu}{\sigma}\right] = 0$,

$$M''(0) = E\left[\left(\frac{X_j-\mu}{\sigma}\right)^2\right] = Var\left(\frac{X_j-\mu}{\sigma}\right) = 1,$$

and

$$M_{Y_n}(s) = \left(M\left(s/\sqrt{n}\right)\right)^n \quad \text{as} \quad Y_n = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \frac{X_j-\mu}{\sigma}.$$

Central Limit Theorem.

Proof (cont.): It is sufficient to show that $\lim_{n\to\infty} M_{Y_n}(s) = \exp\left\{\frac{s^2}{2}\right\}$.

Let $M(s)$ denote the moment generating function of $\frac{X_j-\mu}{\sigma}$. Then,
$M(0)=1$, $M'(0) = E\left[\frac{X_j-\mu}{\sigma}\right] = 0$,

$$M''(0) = E\left[\left(\frac{X_j-\mu}{\sigma}\right)^2\right] = Var\left(\frac{X_j-\mu}{\sigma}\right) = 1,$$

and

$$M_{Y_n}(s) = \left(M\left(s/\sqrt{n}\right)\right)^n \qquad \text{as} \qquad Y_n = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{X_j-\mu}{\sigma}.$$

Indeed,

$$M_{Y_n}(s) = E\left[\exp\left\{\frac{s}{\sqrt{n}}\sum_{j=1}^{n}\frac{X_j-\mu}{\sigma}\right\}\right] = \prod_{j=1}^{n} E\left[\exp\left\{\frac{s}{\sqrt{n}}\frac{X_j-\mu}{\sigma}\right\}\right] = \left(M\left(\frac{s}{\sqrt{n}}\right)\right)^n.$$

Central Limit Theorem.

Proof (cont.): It is sufficient to show that $\lim_{n\to\infty} M_{Y_n}(s) = \exp\left\{\frac{s^2}{2}\right\}$.

Let $M(s)$ denote the moment generating function of $\frac{X_j-\mu}{\sigma}$. Then, $M(0)=1$, $M'(0)=0$, $M''(0)=1$ and $M_{Y_n}(s) = \left(M\left(s/\sqrt{n}\right)\right)^n$. Therefore,

$$\ln M_{Y_n}(s) = n \ln M\left(\frac{s}{\sqrt{n}}\right),$$

where Taylor expansion yields

$$M(h) = M(0) + M'(0)h + \frac{1}{2}M''(0)h^2 + O(h^3) = 1 + \frac{1}{2}h^2 + O(h^3)$$

and $\ln(1+x) = x + O(x^2)$. Thus, for a given $s \in \mathbb{R}$,

$$\ln M_{Y_n}(s) = n \ln M\left(\frac{s}{\sqrt{n}}\right) = n \ln\left(1 + \frac{s^2}{2n} + O\left(\frac{1}{n\sqrt{n}}\right)\right) = n\left(\frac{s^2}{2n} + O\left(\frac{1}{n\sqrt{n}}\right)\right).$$

Hence, $\ln M_{Y_n}(s) = \frac{s^2}{2} + O\left(\frac{1}{\sqrt{n}}\right) \to \frac{s^2}{2}$ and $M_{Y_n}(s) \to \exp\left\{\frac{s^2}{2}\right\}$ as $n \to \infty$.

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

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

Example. Consider a normal random variable X with mean μ and variance σ^2, i.e., $X \sim \mathcal{N}(\mu, \sigma^2)$. Then, $Z = \frac{X-\mu}{\sigma}$ is a standard normal random variable. Thus, the moment generating function of $X = \sigma Z + \mu$ equals

$$M_X(s) = E\left[e^{sX}\right] = e^{s\mu} E\left[e^{s\sigma Z}\right] = e^{s\mu} M_Z(s\sigma) = \exp\{\mu s\} \exp\left\{\frac{\sigma^2 s^2}{2}\right\}$$

$$= \exp\left\{\mu s + \frac{\sigma^2 s^2}{2}\right\}, \qquad s \in \mathbb{R}.$$

One-sided Chebyshev inequality.

Recall Markov and Chebyshev inequalities.

Theorem (Markov inequality). If X is a random variable that takes only nonnegative values with finite mean $E[X] = \mu$, then for any $\alpha > 0$,

$$P(X \geq \alpha) \leq \frac{E[X]}{\alpha} = \frac{\mu}{\alpha}$$

Theorem (Chebyshev inequality). If X is a random variable with finite mean $E[X] = \mu$ and variance $Var(X) = \sigma^2$, then for any $\alpha > 0$,

$$P(|X - \mu| \geq \alpha) \leq \frac{\sigma^2}{\alpha^2}$$

Next, we prove the following inequality.

Theorem (One-sided Chebyshev inequality). If X is a random variable with mean $E[X] = 0$ and variance $Var(X) = \sigma^2$, then for any $\alpha > 0$,

$$P(X \geq \alpha) \leq \frac{\sigma^2}{\sigma^2 + \alpha^2}$$

One-sided Chebyshev inequality.

Theorem (One-sided Chebyshev inequality). If X is a random variable with mean $E[X] = 0$ and variance $Var(X) = \sigma^2$, then for any $\alpha > 0$,

$$P(X \geq \alpha) \;\leq\; \frac{\sigma^2}{\sigma^2 + \alpha^2}$$

Proof: For any $\beta > 0$, by Markov inequality, we have

$$P(X \geq \alpha) = P(X+\beta \geq \alpha+\beta) \leq P\Big((X+\beta)^2 \geq (\alpha+\beta)^2\Big) \;\leq\; \frac{E\big[(X+\beta)^2\big]}{(\alpha+\beta)^2}$$

Hence, since $\sigma^2 = Var(X) = E[X^2] - \big(E[X]\big)^2 = E[X^2]$

$$P(X \geq \alpha) \leq \frac{E\big[(X+\beta)^2\big]}{(\alpha+\beta)^2} = \frac{E[X^2] + 2\beta E[X] + \beta^2}{(\alpha+\beta)^2} = \frac{\sigma^2+\beta^2}{(\alpha+\beta)^2} = \varphi(\beta).$$

for any β in the domain $(0, \infty)$. Function $\varphi(\beta) = \frac{\sigma^2+\beta^2}{(\alpha+\beta)^2}$ is minimized at $\beta = \frac{\sigma^2}{\alpha}$, yielding the following upper bound

$$P(X \geq \alpha) \leq \varphi\left(\frac{\sigma^2}{\alpha}\right) = \frac{\sigma^2}{\sigma^2+\alpha^2}.$$

One-sided Chebyshev inequality.

Theorem (Chebyshev inequality). If X is a random variable with finite mean $E[X] = \mu$ and variance $Var(X) = \sigma^2$, then for any $\alpha > 0$,

$$P(|X - \mu| \geq \alpha) \ \leq \ \frac{\sigma^2}{\alpha^2}$$

Theorem (One-sided Chebyshev inequality). If X is a random variable with mean $E[X] = 0$ and variance $Var(X) = \sigma^2$, then for any $\alpha > 0$,

$$P(X \geq \alpha) \ \leq \ \frac{\sigma^2}{\sigma^2 + \alpha^2}$$

Corollary. If X is a random variable with mean $E[X] = \mu$ and variance $Var(X) = \sigma^2$, then for any $\alpha > 0$,

$$P(X \geq \mu + \alpha) = P(X - \mu \geq \alpha) \ \leq \ \frac{\sigma^2}{\sigma^2 + \alpha^2}$$

Chernoff bound.

Theorem (Chernoff bound). If X is a random variable with finite moment generating function $M_X(s)$, then for any $\alpha \in \mathbb{R}$,

$$P(X \geq \alpha) \;\leq\; e^{-s\alpha} M_X(s)$$

for all $s > 0$ in the domain of $M_X(s)$.

Proof. For any given $s > 0$ in the domain of $M_X(s)$, Markov inequality yields

$$P(X \geq \alpha) = P(sX \geq s\alpha) = P\left(e^{sX} \geq e^{s\alpha}\right) \leq \frac{E\left[e^{sX}\right]}{e^{s\alpha}} = e^{-s\alpha} M_X(s)$$

Example. Consider a standard normal random variable Z and $a > 0$. Then,

$$P(Z \geq a) \leq e^{-sa} M_X(s) = e^{-sa} e^{s^2/2} = e^{(s^2-2sa)/2} = e^{-a^2/2} e^{(s-a)^2/2},$$

where the right hand side $e^{-a^2/2} e^{(s-a)^2/2}$ is minimized when $s = a$, yielding

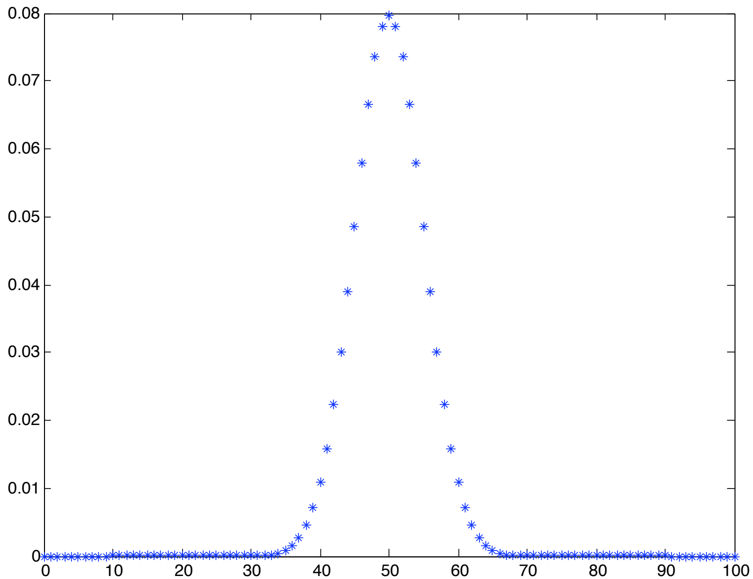
$$P(Z \geq a) \leq e^{-a^2/2}.$$

Chernoff bound.

Example. Consider a binomial random variable S_n with parameters (n, p).

$$S_n = E[S_n] \pm \sqrt{Var(S_n)} = np \pm \sqrt{np(1-p)}.$$

For $a \in (p, 1)$, we want to find an upper bound on the tail probability $P(S_n \geq an)$.



Case $n = 100, \ p = \frac{1}{2}$.

Here, $S_n = 50 \pm 5$.

Chernoff bound.

Example (cont.): Consider a binomial random variable S_n with parameters (n, p). For $a \in (p, 1)$, we want to find an upper bound on the tail probability $P(S_n \geq an)$. Applying Chernoff bound, we have

$$P(S_n \geq an) \leq e^{-san} M_{S_n}(s) = e^{-san}\left(1 - p + pe^s\right)^n = \exp\left\{-n\left(sa - \ln(1 - p + pe^s)\right)\right\}$$

for all $s > 0$. We minimize the right-hand side:

$$\frac{d}{ds}\left(sa - \ln(1 - p + pe^s)\right) = a - \frac{pe^s}{1 - p + pe^s} = 0$$

yielding the extremum at $s^* = \ln\left(\frac{a(1-p)}{p(1-a)}\right) > 0$. Plugging s^* into the above Chernoff bound, we obtain

$$P(S_n \geq an) \leq \exp\left\{-n\left(s^* a - \ln(1 - p + pe^{s^*})\right)\right\} = \exp\{-nH(a|p)\},$$

$$\text{where} \qquad H(a|p) = -a \ln\left(\frac{p}{a}\right) - (1 - a)\ln\left(\frac{1-p}{1-a}\right)$$

is **relative entropy** (aka Kullback-Leibler divergence).

Law of Large Numbers (LLN):

$$\frac{S_n}{n} \xrightarrow{d} p \quad \text{as } n \to \infty \quad \Leftrightarrow \quad S_n = np + \text{`smaller terms'}$$

Central Limit Theorem (CLT):

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{w} Z \text{ as } n \to \infty \quad \Leftrightarrow \quad S_n = np + Z\sqrt{np(1-p)} + \text{`smaller terms'},$$

where $Z \sim \mathcal{N}(0,1)$ (standard normal), and

$$P(S_n \geq np + b\sqrt{n}) = P\Big(Z\sqrt{np(1-p)} \geq b\sqrt{n}\Big) + o(1) \longrightarrow P\Big(Z \geq \frac{b}{\sqrt{p(1-p)}}\Big)$$

as $n \to \infty$, while

$$P(S_n \geq an) = P\Big(np + Z\sqrt{np(1-p)} \geq an\Big) + o(1) = P\Big(Z \geq \frac{a-p}{\sqrt{p(1-p)}}\sqrt{n}\Big) + o(1).$$

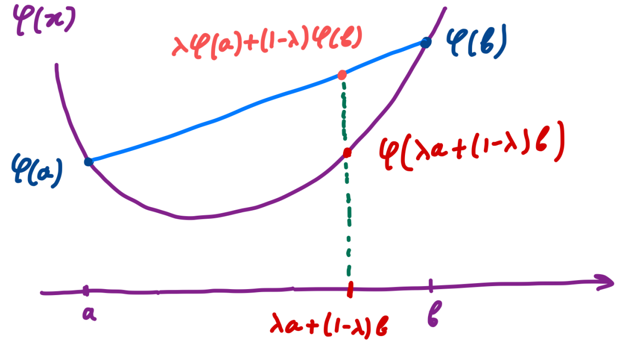
Hence, $P(S_n \geq an) \longrightarrow 0$ as $n \to \infty$.

Chernoff bound: $P(S_n \geq an) \leq \exp\{-nH(a|p)\}$

Large Deviations Theory:

$$P(S_n \geq an) \asymp \exp\{-nH(a|p)\} \quad \Leftrightarrow \quad \lim_{n\to\infty} \frac{-1}{n} \ln P(S_n \geq an) = H(a|p).$$

Convex functions.



A function $\varphi(x)$ is said to be convex over an interval $\mathcal{I}$, the domain of the function, if

$$\varphi(\lambda a + (1-\lambda)b) \;\leq\; \lambda\varphi(a) + (1-\lambda)\varphi(b)$$

for all $\lambda \in [0,1]$ and all real a and b in $\mathcal{I}$.

If function $\varphi(x)$ is twice differentiable, then

$$\varphi(x) \text{ is convex in } \mathcal{I} \quad \Leftrightarrow \quad \varphi''(x) \geq 0 \quad \forall x \in \mathcal{I}$$

A function $\varphi(x)$ is said to be concave if $-\varphi(x)$ is convex. If function $\varphi(x)$ is twice differentiable, then

$$\varphi(x) \text{ is concave in } \mathcal{I} \quad \Leftrightarrow \quad \varphi''(x) \leq 0 \quad \forall x \in \mathcal{I}$$

Jensen's inequality.

A function $\varphi(x)$ is said to be convex over an interval $\mathcal{I}$, the domain of the function, if

$$\varphi(\lambda a + (1-\lambda)b) \;\le\; \lambda\varphi(a) + (1-\lambda)\varphi(b)$$

for all $\lambda \in [0,1]$ and all real a and b in $\mathcal{I}$.

Jensen's inequality: Suppose φ is convex. Then

$$\varphi(E[X]) \le E[\varphi(X)]$$

Proof. Let $\mu = E[X]$. There is a line $\ell(x) = ax + b$ such that

$$\ell(x) \le \varphi(x) \quad \text{and} \quad \ell(\mu) = \varphi(\mu)$$

Then

$$\varphi(\mu) = \ell(\mu) = E[\ell(X)] \le E[\varphi(X)]$$

□

Jensen's inequality.

Jensen's inequality: Suppose φ is convex. Then

$$\varphi(E[X]) \leq E[\varphi(X)]$$

Examples:

- $E[X^2] \geq \big(E[X]\big)^2$ as $\varphi(x) = x^2$ is convex for $x \in \mathbb{R}$.

- For any given $a \in \mathbb{R}$, $E[e^{aX}] \geq e^{aE[X]}$ as $\varphi(x) = e^{ax}$ is convex for $x \in \mathbb{R}$.

- If $X \geq 0$ then $E[X^3] \geq \big(E[X]\big)^3$ as $\varphi(x) = x^3$ is convex for $x \in [0, \infty)$.

- If $X > 0$ then $E[X \cdot \ln(X)] \geq E[X] \cdot \ln(E[X])$ as $\varphi(x) = x\ln(x)$ is convex for $x \in (0, \infty)$.

- If $X > 0$ then $E[\ln(X)] \leq \ln(E[X])$ as $\varphi(x) = \ln(x)$ is concave for $x \in (0, \infty)$.

Characteristic function.

Definition. The characteristic function $\varphi_X : \mathbb{R} \to \mathbb{C}$ of a random variable X is defined by

$$\varphi_X(s) = E\left[e^{isX}\right] \qquad \forall s \in \mathbb{R}.$$

Properties:

- Euler's formula states that $e^{i\theta} = \cos\theta + i\sin\theta$ for all $\theta \in \mathbb{R}$.

Therefore,

$$\varphi_X(s) = E\left[e^{isX}\right] = E\left[\cos(sX)\right] + i\,E\left[\sin(sX)\right]$$

is well defined for all $s \in \mathbb{R}$.

- $$\varphi_X(s) = E\left[e^{isX}\right] = \begin{cases} \displaystyle\sum_{x:\, p_X(x)>0} e^{isx} p_X(x) & \text{if } X \text{ is discrete,} \\ \displaystyle\int_{-\infty}^{\infty} e^{isx} f_X(x)\, dx & \text{if } X \text{ is continuous.} \end{cases}$$

Characteristic function.

Definition. The characteristic function $\varphi_X : \mathbb{R} \to \mathbb{C}$ of a random variable X is defined by

$$\varphi_X(s) = E\left[e^{isX}\right] \qquad \forall s \in \mathbb{R}.$$

- $\varphi_X(0) = 1$.

- The derivatives of $\varphi_X(s)$ are computed as follows

$$\varphi_X^{(n)}(s) = \frac{d^n}{ds^n} E[e^{isX}] = E\left[\frac{d^n}{ds^n} e^{isX}\right] = i^n\, E[X^n e^{isX}].$$

Thus, $E[X^n] = (-i)^n \varphi_X^{(n)}(0)$ (the n^{th} moment) as $-i = \frac{1}{i}$.

- If $X_1, X_2, \ldots, X_n$ are independent random variables, then the characteristic function of $X = X_1 + X_2 + \ldots + X_n$ equals

$$\varphi_X(s) = \varphi_{X_1}(s) \cdot \varphi_{X_2}(s) \cdot \ldots \cdot \varphi_{X_n}(s).$$

- CLT can be proved via characteristic functions without assuming the moment generating function of X_j is well defined.

Characteristic function.

Definition. The characteristic function $\varphi_X : \mathbb{R} \to \mathbb{C}$ of a random variable X is defined by

$$\varphi_X(s) = E\left[e^{isX}\right] \qquad \forall s \in \mathbb{R}.$$

Connection to harmonic analysis: for a continuous random variable,

$$\frac{1}{\sqrt{2\pi}}\varphi_X(-s) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-isx} f_{\mathsf{x}}(x)\,dx$$

is a Fourier transform of $f_{\mathsf{x}}(x)$, and

$$\frac{1}{\sqrt{2\pi}}\varphi_X(s) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{isx} f_{\mathsf{x}}(x)\,dx$$

is the inverse Fourier transform of $f_{\mathsf{x}}(x)$.

Similar statements apply in the case of a discrete random variable.

Generating function.

Definition. For a given random variable X, the function

$$G_X(s) = E\Big[s^X\Big], \quad s > 0,$$

is called the generating function.

• Connection to m.g.f. $G_X(s) = M_X\Big(\ln s\Big)$.

• $$G_X(s) = E\Big[s^X\Big] = \begin{cases} \displaystyle\sum_{x:\, p_X(x)>0} s^x\, p_X(x) & \text{if } X \text{ is discrete,} \\ \displaystyle\int_{-\infty}^{\infty} s^x\, f_X(x)\, dx & \text{if } X \text{ is continuous.} \end{cases}$$

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

$$G_X(s) = G_{X_1}(s) \cdot G_{X_2}(s) \cdot \ldots \cdot G_{X_n}(s).$$

• If X is nonnegative integer valued random variable, i.e., $X = 0, 1, 2, \ldots$, then its generating function is convex.