MTH 464/564

Lectures 25-28

Yevgeniy Kovchegov

Oregon State University
• Jensen’s inequality.

• Characteristic and generating functions.

• Branching processes.

• Size biasing.

• Functions of random variables.
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

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

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

$$\varphi(x) \text{ is concave in } \mathcal{I} \iff \varphi''(x) \leq 0 \quad \forall x \in \mathcal{I}$$
**Jensen’s inequality.**

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

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

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

**Jensen's inequality:** Suppose $\varphi$ is convex. Then

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

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

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

Then

$$\varphi(\mu) = \ell(\mu) = E[\ell(X)] \leq E[\varphi(X)]$$
Jensen’s inequality.

**Jensen’s inequality:** Suppose $\varphi$ is convex. Then

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

**Examples:**

- $E[X^2] \geq (E[X])^2$ 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 $E[X^3] \geq (E[X])^3$ 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 $\varphi_X : \mathbb{R} \rightarrow \mathbb{C}$ of a random variable $X$ is defined by

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

Properties:

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

Therefore,

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

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

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

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

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

- $\varphi_X(0) = 1$.

- 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

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

Connection to harmonic analysis: for a continuous random variable,

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

is a Fourier transform of $f_X(x)$, and

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

is the inverse Fourier transform of $f_X(x)$.

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

Definition. For a given random variable $X$, the function

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

is called the generating function.

- **Connection to m.g.f.** $G_X(s) = M_X\left(\ln s\right)$.

- $G_X(s) = E\left[s^X\right] = \begin{cases} \sum_{x:p_x(x)>0} s^x p_x(x) & \text{if } X \text{ is discrete,} \\ \int_{-\infty}^{\infty} s^x f_x(x) \, dx & \text{if } X \text{ is continuous.} \end{cases}$

- 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.
Branching process.

Problem of Extinction. Start in the 0th generation with 1 parent. In the first generation we shall have 0, 1, 2, ... offsprings with respective probabilities

\[ p_0, p_1, p_2, \ldots \]

If in the \( t \)th generation there are \( Z_t = k \) individuals, then in the \((t + 1)\)st generation there will be

\[ Z_{t+1} = X_1 + X_2 + \ldots + X_k \] offsprings,

where \( X_1, X_2, \ldots, X_k \) are independent random variables, each with the same probability mass function \( p_0, p_1, p_2, \ldots \).

Question: For which probability mass functions \( \{p_k\} \) do we have guaranteed extinction of the genealogical (family) tree?

Note that case \( p_1 = 1 \) is trivial.

Solution: For \( m = 0, 1, \ldots \), let

\[ A_m = \{\text{extinction by } m\text{th generation}\} \]

Then, \( d_m = P(A_m) \) is the probability that the process dies out by the \( m \)th generation.
Branching process.

Solution (cont.): For \( m = 0, 1, \ldots \), let

\[ A_m = \{ \text{extinction by } m^{\text{th}} \text{ generation} \}. \]

Then, \( d_m = P(A_m) \) is the probability that the process dies out by the \( m \)th generation. Since, \( A_m \subseteq A_{m+1} \),

\[ 0 = d_0 \leq d_1 \leq d_2 \leq \ldots \leq 1 \]

and \( \lim_{m \to \infty} d_m \) exists. Observe that \( \bigcup_{m=1}^{\infty} A_m = \{ \text{extinction} \} \) and

\[ d = P \left( \bigcup_{m=1}^{\infty} A_m \right) = \lim_{m \to \infty} P(A_m) = \lim_{m \to \infty} d_m \]

is the probability of extinction.

Next, observe that for \( m \geq 1 \),

\[ d_m = P(A_m) = \sum_{k=0}^{\infty} P(Z_1 = k) P\left( A_m \mid Z_1 = k \right) = \sum_{k=0}^{\infty} p_k (d_{m-1})^k. \]
Branching process.

Solution (cont.): Since, \( A_m \subseteq A_{m+1} \),
\[
0 = d_0 \leq d_1 \leq d_2 \leq \ldots \leq 1
\]
and \( \lim_{m \to \infty} d_m \) exists. Observe that
\[
\bigcup_{m=1}^{\infty} A_m = \{ \text{extinction} \}
\]
and
\[
d = P\left( \bigcup_{m=1}^{\infty} A_m \right) = \lim_{m \to \infty} P(A_m) = \lim_{m \to \infty} d_m
\]
is the probability of extinction.

Next, observe that for \( m \geq 1 \),
\[
d_m = P(A_m) = \sum_{k=0}^{\infty} P(Z_1 = k) P\left( A_m \mid Z_1 = k \right) = \sum_{k=0}^{\infty} p_k (d_{m-1})^k.
\]
Let \( h(z) \) be the generating function for the distribution \( p_k \):
\[
h(z) = \sum_{k=0}^{\infty} p_k z^k.
\]
Then,
\[
d_m = h(d_{m-1}), \quad \text{and as } d_m \to d, \quad \text{we have } d = h(d).
\]
Branching process.

Figure 10.2: Graphs of \( y = z \) and \( y = h(z) \).

\( y = h(z) \) can intersect the line \( y = z \) in at most two points. Since we know it must intersect the line \( y = z \) at \((1, 1)\), we know that there are just three possibilities, as shown in Figure 10.2.

In case (a) the equation \( d = h(d) \) has roots \( \{d, 1\} \) with \( 0 \leq d < 1 \). In the second case (b) it has only the one root \( d = 1 \). In case (c) it has two roots \( \{1, d\} \) where \( 1 < d \).

Since we are looking for a solution \( 0 \leq d \leq 1 \), we see in cases (b) and (c) that our only solution is \( 1 \). In these cases we can conclude that the process will die out with probability 1.

However in case (a) we are in doubt. We must study this case more carefully.

From Equation 10.4 we see that

\[
 h_0(1) = p_1 + 2p_2 + 3p_3 + \cdots = m,
\]

where \( m \) is the expected number of offspring produced by a single parent.

In case (a) we have \( h_0(1) > 1 \), in (b) \( h_0(1) = 1 \), and in (c) \( h_0(1) < 1 \).

Thus our three cases correspond to \( m > 1 \), \( m = 1 \), and \( m < 1 \).

We assume now that \( m > 1 \). Recall that \( d_0 = 0, d_1 = h(d_0) = p_0, d_2 = h(d_1), \ldots, d_n = h(d_{n-1}) \).

We can construct these values geometrically, as shown in Figure 10.3.

We can see geometrically, as indicated for \( d_0, d_1, d_2, \) and \( d_3 \) in Figure 10.3, that the points \((d_i, h(d_i))\) will always lie above the line \( y = z \).

Hence, they must converge to the first intersection of the curves \( y = z \) and \( y = h(z) \) (i.e., to the root \( d < 1 \)).

This leads us to the following theorem.

**Theorem 10.2**

Consider a branching process with generating function \( h(z) \) for the number of offspring of a given parent. Let \( d \) be the smallest root of the equation \( z = h(z) \).

If the mean number \( m \) of offspring produced by a single parent is \( \leq 1 \), then \( d = 1 \) and the process dies out with probability 1. If \( m > 1 \) then \( d < 1 \) and the process dies out with probability \( d \).

Source: *Grinstead and Snell (Chapter 10)*

\[
 d_m = h(d_{m-1}), \quad \text{and as } d_m \to d, \quad \text{we have } d = h(d).
\]
Branching process.

Solution (cont.):

\[ d_m = h(d_{m-1}), \quad \text{and as } d_m \to d, \quad \text{we have } d = h(d). \]

- \( h(z) = \sum_{k=0}^{\infty} p_k z^k \), its derivative \( h'(z) = \sum_{k=1}^{\infty} k p_k z^{k-1} \), and
  \[ h'(1) = \sum_{k=1}^{\infty} k p_k = E[X_i]. \]

- \( h(z) \) is a convex function as
  \[ h''(z) = \sum_{k=2}^{\infty} k(k-1) p_k z^{k-2} \geq 0, \quad (z \geq 0). \]

**Extinction criterium:** Suppose \( p_1 \neq 1 \). Then,
\[ d = 1 \quad (\text{guaranteed extinction}) \quad \text{if and only if} \quad h'(1) = E[X_i] \leq 1. \]
Branching process.

$h(z)$ is a convex function for $z \geq 0$ as $h''(z) = 2p_2 + 6p_3z + \ldots \geq 0$

d = 1 \quad \text{if and only if} \quad h'(1) = E[X_i] \leq 1.
Branching process.

\[ d_m = h(d_{m-1}) \quad \text{and} \quad d = h(d). \]
Critical branching process.

Example. Consider a critical binary Galton-Watson (branching) process:

\[ p_0 = p_2 = \frac{1}{2} \]

It is critical: \( E[X_i] = p_1 + 2p_2 + 3p_3 + \ldots = 1 \).

Let \( N \) be the number vertices. Then,

\[ P(N < \infty) = 1 \quad \text{and} \quad E[N] = \infty \]

Example. Consider a Galton-Watson (branching) process with \( p_0 = \frac{1}{2}, \; p_1 = \frac{1}{4}, \; p_2 = \frac{1}{8}, \; \ldots, \; p_k = \frac{1}{2^{k+1}}, \; \ldots \).

It is critical: \( E[X_i] = p_1 + 2p_2 + 3p_3 + \ldots = 1 \).

Here too, for the number of vertices \( N \),

\[ P(N < \infty) = 1 \quad \text{and} \quad E[N] = \infty \]
Size biasing.

Jensen’s inequality: If $\varphi$ is convex, then $\varphi(E[X]) \leq E[\varphi(X)]$.

Suppose $X$ is a positive valued continuous random variable ($X > 0$) with mean $\mu > 0$, variance $\sigma^2 > 0$ and probability density function $f_x(x)$. By Jensen’s inequality we have a lower bound

$$E[X \cdot \ln X] \geq E[X] \cdot \ln(E[X]) = \mu \ln \mu$$

as $\varphi(x) = x \ln x$ is convex for $x \in (0, \infty)$.

**Problem:** Find an upper bound on $E[X \cdot \ln X]$.

**Size biasing:** Function $g(x) = \frac{1}{\mu} x f_x(x)$ is a probability density function as

$$\int_0^\infty g(x) \, dx = \frac{1}{\mu} \int_0^\infty x f_x(x) \, dx = \frac{1}{\mu} E[X] = 1.$$

Let $Y$ be a random variable with p.d.f. $g(x)$, then since $\ln x$ is concave,

$$E[X \cdot \ln X] = \int_0^\infty x \ln x \cdot f_x(x) \, dx = \mu \int_0^\infty \ln x \cdot g(x) \, dx = \mu E[\ln Y] \leq \mu \ln(E[Y]) = \mu \ln \left( \mu + \frac{\sigma^2}{\mu} \right)$$

by Jensen’s inequality, where $E[Y] = \int_0^\infty x g(x) \, dx = \frac{1}{\mu} \int_0^\infty x^2 f_x(x) \, dx = \frac{E[X^2]}{\mu} = \frac{\sigma^2 + \mu^2}{\mu}$. 
Size biasing.

Suppose $X$ is a positive valued continuous random variable ($X > 0$) with mean $\mu > 0$, variance $\sigma^2 > 0$ and probability density function $f_x(x)$.

**Jensen’s inequality:** a lower bound

$$E[X \cdot \ln X] \geq \mu \ln \mu$$

**Size biasing:** an upper bound

$$E[X \cdot \ln X] \leq \mu \ln \left( \mu + \frac{\sigma^2}{\mu} \right)$$

Hence,

$$\mu \ln \mu \leq E[X \cdot \ln X] \leq \mu \ln \left( \mu + \frac{\sigma^2}{\mu} \right)$$

The inequalities hold if $X$ is a positive valued discrete random variable.

**Example.** Let $X$ be an exponential random variable with parameter $\lambda > 0$, then $\mu = \sigma = \lambda$ and

$$\lambda \ln \lambda \leq E[X \cdot \ln X] \leq \lambda \ln \left( 2\lambda \right) = \lambda \left( \ln \lambda + \ln 2 \right)$$
Functions of random variables.

**Theorem.** Let $X$ be a continuous random variable with density function $f_x(x)$. If $g(x)$ is a strictly monotone (increasing or decreasing) differentiable function, and if $Y = g(X)$, then the probability density function of $Y$

$$f_y(y) = \begin{cases} f_x(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right| & \text{if } y = g(x) \text{ for some } x \text{ s.t. } f_x(x) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

where $g^{-1}$ is the inverse of $g$: $g(x) = y \iff g^{-1}(y) = x$.

**Question:** Let $X_1$ and $X_2$ be continuous random variable with the joint probability density function $f_{x_1,x_2}(x_1, x_2)$. Let $g(x_1, x_2) = \left( g_1(x_1, x_2), g_2(x_1, x_2) \right)$ be a bijection (one-to-one and onto) mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$. Find the joint probability density function $f_{y_1,y_2}(y_1, y_2)$ of $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$. 
Functions of random variables.

**Question:** Let $X_1$ and $X_2$ be continuous random variable with the joint probability density function $f_{x_1,x_2}(x_1, x_2)$. Let

$$g(x_1, x_2) = \left(g_1(x_1, x_2), g_2(x_1, x_2)\right)$$

be a bijection (one-to-one and onto) mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$. Find the joint probability density function $f_{y_1,y_2}(y_1, y_2)$ of

$$Y_1 = g_1(X_1, X_2) \quad \text{and} \quad Y_2 = g_2(X_1, X_2).$$

**Theorem.**

$$f_{y_1,y_2}(y_1, y_2) = f_{x_1,x_2}(x_1, x_2) \cdot \left|\frac{\partial g(x_1, x_2)}{\partial x_1 \partial x_2}\right|^{-1}, \quad \text{where } (x_1, x_2) = g^{-1}(y_1, y_2)$$

if $f_{x_1,x_2}(g^{-1}(y_1, y_2)) = f_{x_1,x_2}(x_1, x_2) > 0$. Here,

$$\frac{\partial g(x_1, x_2)}{\partial x_1 \partial x_2} = \det\left(\begin{array}{cc} \frac{\partial g_1(x_1, x_2)}{\partial x_1} & \frac{\partial g_1(x_1, x_2)}{\partial x_2} \\ \frac{\partial g_2(x_1, x_2)}{\partial x_1} & \frac{\partial g_2(x_1, x_2)}{\partial x_2} \end{array}\right)$$

is the Jacobian of $g(x_1, x_2)$. 
Functions of random variables.

Theorem.

\[ f_{y_1,y_2}(y_1, y_2) = f_{x_1,x_2}(x_1, x_2) \cdot \left| \frac{\partial g(x_1, x_2)}{\partial x_1 \partial x_2} \right|^{-1} \], \quad \text{where } (x_1, x_2) = g^{-1}(y_1, y_2)

if \( f_{x_1,x_2}(g^{-1}(y_1, y_2)) = f_{x_1,x_2}(x_1, x_2) > 0 \). Here,

\[
\frac{\partial g(x_1, x_2)}{\partial x_1 \partial x_2} = \det \left( \begin{array}{cc}
\frac{\partial g_1(x_1, x_2)}{\partial x_1} & \frac{\partial g_1(x_1, x_2)}{\partial x_2} \\
\frac{\partial g_2(x_1, x_2)}{\partial x_1} & \frac{\partial g_2(x_1, x_2)}{\partial x_2}
\end{array} \right)
\]

is the Jacobian of \( g(x_1, x_2) \).

Example. Let \( X_1 \) be an exponential random variable with parameter \( \lambda_1 = 1 \) and \( X_2 \) be an exponential random variable with parameter \( \lambda_2 = 2 \). Suppose \( X_1 \) and \( X_2 \) are independent. Find the joint probability density function \( f_{y_1,y_2}(y_1, y_2) \) of

\[ \begin{align*}
Y_1 &= X_1 + X_2 \quad \text{and} \quad Y_2 = \frac{X_1}{X_1 + X_2}.
\end{align*} \]