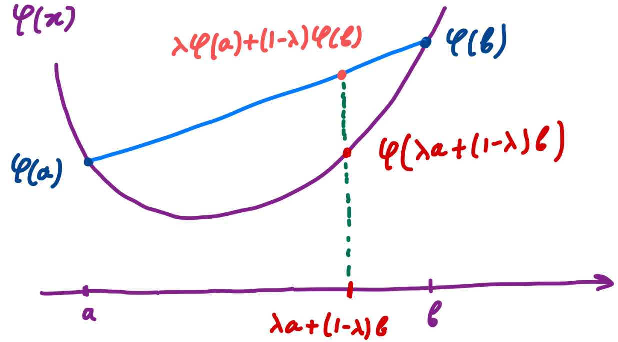
MTH 464/564

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

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

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

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

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

Jensen's inequality: Suppose φ 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

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

$\square$

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

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

is called the generating function.

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

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

• 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 $0, 1, 2, \ldots$ 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 $(t+1)^{\text{st}}$ generation there will be

$$Z_{t+1} = X_1 + X_2 + \ldots + X_k \quad \text{offsprings},$$

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

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

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

$$0 = d_0 \le d_1 \le d_2 \le \ldots \le 1$$

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 $m \ge 1$,

$$d_m = P(A_m) = \sum_{k=0}^{\infty} P(Z_1 = k)\, P\Big(A_m \,\Big|\, Z_1 = k\Big) = \sum_{k=0}^{\infty} p_k\, (d_{m-1})^k.$$

Branching process.

Solution (cont.): Since, $A_m \subseteq A_{m+1}$,

$$0 = d_0 \le d_1 \le d_2 \le \ldots \le 1$$

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 $m \ge 1$,

$$d_m = P(A_m) = \sum_{k=0}^{\infty} P(Z_1 = k)\, P\Big(A_m \,\Big|\, Z_1 = k\Big) = \sum_{k=0}^{\infty} p_k\,(d_{m-1})^k.$$

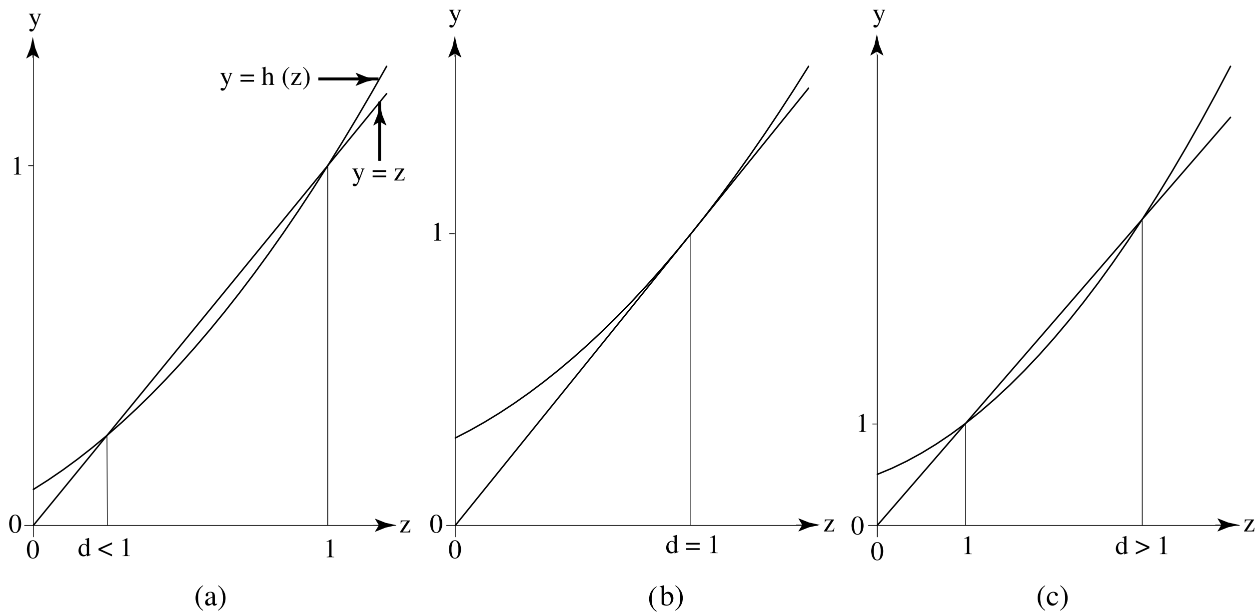
Let $h(z)$ be the generating function for the distribution p_k:

$$h(z) = \sum_{k=0}^{\infty} p_k\, z^k.$$

Then,

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

Branching process.



Source: *Grinstead and Snell (Chapter 10)*

$d_m = h(d_{m-1})$, and as $d_m \to d$, 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

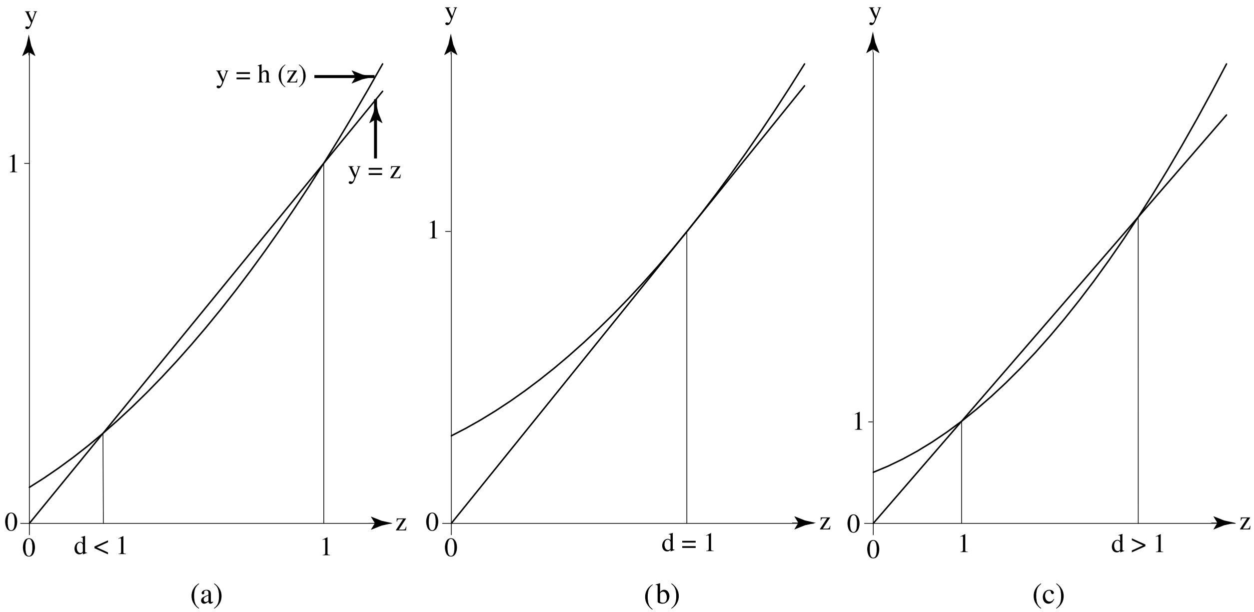
$$h''(z) = \sum_{k=2}^{\infty} k(k-1)\, p_k\, z^{k-2} \geq 0, \qquad (z \geq 0).$$

Extinction criterium: Suppose $p_1 \neq 1$. Then,

$d = 1$ (guaranteed extinction) if and only if $h'(1) = E[X_i] \leq 1$.

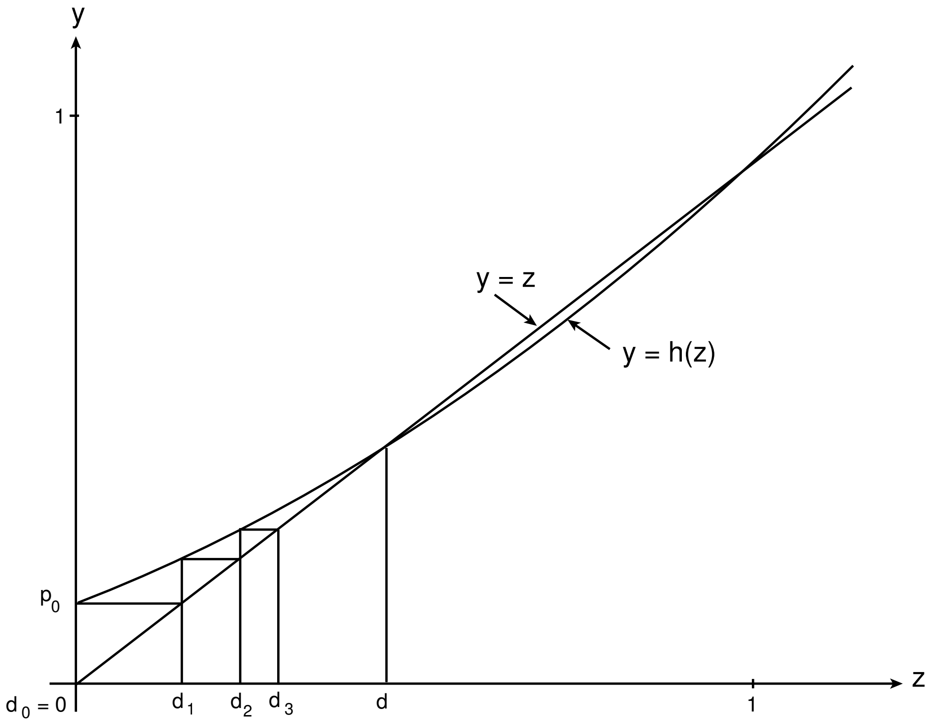
Branching process.

$h(z)$ is a convex function for $z \geq 0$ as $h''(z) = 2p_2 + 6p_3 z + \ldots \geq 0$



$d = 1$ if and only if $h'(1) = E[X_i] \leq 1$.

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 $p_0 = \frac{1}{2}, \quad p_1 = \frac{1}{4}, \quad p_2 = \frac{1}{8}, \quad \ldots, \quad p_k = \frac{1}{2^{k+1}}, \quad \ldots.$
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 φ is convex, then $\varphi(E[X]) \le E[\varphi(X)]$.

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

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

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 $g(x) = \frac{1}{\mu} x f_{\mathsf{x}}(x)$ is a probability density function as $\int_0^\infty g(x)\,dx = \frac{1}{\mu}\int_0^\infty x f_{\mathsf{x}}(x)\,dx = \frac{1}{\mu}E[X] = 1$.

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

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

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

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 $f_{\mathsf{x}}(x)$. If $g(x)$ is a strictly monotone (increasing or decreasing) differentiable function, and if $Y = g(X)$, then the probability density function of Y

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

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

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

$$g(x_1, x_2) = \Big(g_1(x_1, x_2),\, g_2(x_1, x_2)\Big)$$

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

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

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

$$g(x_1, x_2) = \Big(g_1(x_1, x_2),\ g_2(x_1, x_2)\Big)$$

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 $f_{X_1,X_2}\Big(g^{-1}(y_1, y_2)\Big) = f_{X_1,X_2}(x_1, x_2) > 0$. Here,

$$\frac{\partial g(x_1, x_2)}{\partial x_1 \partial x_2} = \det \begin{pmatrix} \frac{\partial g_1(x_1,x_2)}{\partial x_1} & \frac{\partial g_1(x_1,x_2)}{\partial x_2} \\ \frac{\partial g_2(x_1,x_2)}{\partial x_1} & \frac{\partial g_2(x_1,x_2)}{\partial x_2} \end{pmatrix}$$

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

Functions of random variables.

Theorem.

$$f_{\mathsf{Y}_1,\mathsf{Y}_2}(y_1,y_2) = f_{\mathsf{X}_1,\mathsf{X}_2}(x_1,x_2) \cdot \left|\frac{\partial g(x_1,x_2)}{\partial x_1 \partial x_2}\right|^{-1}, \quad \text{where} \quad (x_1,x_2) = g^{-1}(y_1,y_2)$$

if $f_{\mathsf{X}_1,\mathsf{X}_2}\big(g^{-1}(y_1,y_2)\big) = f_{\mathsf{X}_1,\mathsf{X}_2}(x_1,x_2) > 0$. Here,

$$\frac{\partial g(x_1,x_2)}{\partial x_1 \partial x_2} = \det\begin{pmatrix} \frac{\partial g_1(x_1,x_2)}{\partial x_1} & \frac{\partial g_1(x_1,x_2)}{\partial x_2} \\ \frac{\partial g_2(x_1,x_2)}{\partial x_1} & \frac{\partial g_2(x_1,x_2)}{\partial x_2} \end{pmatrix}$$

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 $f_{\mathsf{Y}_1,\mathsf{Y}_2}(y_1,y_2)$ of

$$Y_1 = X_1 + X_2 \quad \text{and} \quad Y_2 = \frac{X_1}{X_1+X_2}.$$