CIS6930/4930

Random Variables: General Case

Prof. Ye Xia

Continuous Random Variables

- A random variable X is called **continuous** if its distribution function can be expressed as

  $$F(x) = \int_{-\infty}^{x} f(u)du$$

  for some integrable function $f : \mathbb{R} \to [0, \infty)$.

  $f(x)$ is called the **probability density function**.

  Here, integrability is in the usual sense, i.e., Riemann-integrability.

- The distribution function of a continuous random variable is continuous.

Probability Density Function (pdf)

If X has a density function (i.e., X is a continuous random variable), then

- (i) $\int_{-\infty}^{\infty} f(x)dx = 1$;
- (ii) $P(X = x) = 0$ for all $x \in \mathbb{R}$;
- (iii) $P(a \le X \le b) = \int_a^b f(x)dx$.
- More generally, for $B \in \mathcal{R}$ that is sufficiently nice (so that the integral is well-defined), $P(X \in B) = \int_B f(x)dx$.

  (Recall $\mathcal{R}$ is the Borel σ-field on $\mathbb{R}$.)

Independence

- **Definition**: Two random variables X and Y are **independent** if the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent for all $x, y \in \mathbb{R}$.
- This definition holds for all types of random variables.
- If X and Y are independent, so are $g(X)$ and $h(Y)$ where g and h are measurable functions from $\mathbb{R} \to \mathbb{R}$.

Expectation

- There are two ways of calculating the expectation of X.

  $E[X] = \int_\Omega X(\omega)dP = \int_\mathbb{R} xdF(x)$.

  The first one is the definition of $E[X]$.

  $$E[X] = \int_\mathbb{R} xdF(x) = \begin{cases} \int_{-\infty}^{\infty} xf(x)dx & \text{if } X \text{ is continuous} \\ \sum_x xP(X=x) & \text{if } X \text{ is discrete,} \end{cases}$$

  provided $\int_{-\infty}^{\infty} |x|f(x)dx$ or $\sum_x |x|P(X=x)$ is finite; in other words, $E|X| < \infty$.

- For the general case, $\int_\mathbb{R} xdF(x)$ is a Riemann-Steiltjes integral, which is a generalization of the Riemann integral.

- For a mixed random with a finite number of discontinuities in $F(x)$, say at $a_1, a_2, \ldots, a_n$, $E[X]$ can be computed by adding the discrete part and the continuous part. For convenience, let $a_0 = -\infty$ and $a_{n+1} = \infty$.

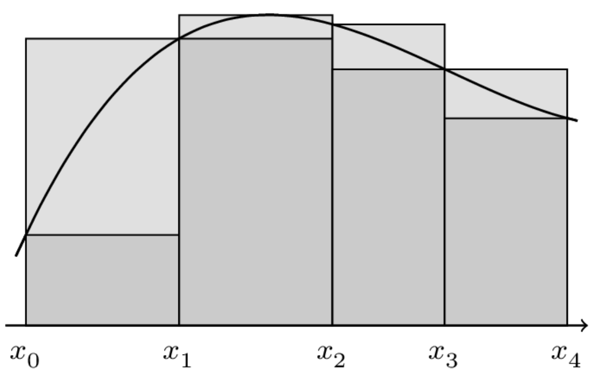
$$E[X] = \sum_{i=1}^{n} a_i P(X = a_i) + \sum_{i=0}^{n} \int_{a_i}^{a_{i+1}} x f(x) dx,$$

where $f(x)$ is the derivative of $F(x)$ on each interval, provided $\int_{a_0}^{a_1} |x| f(x) dx < \infty$ and $\int_{a_n}^{a_{n+1}} |x| f(x) dx < \infty$ (only the integrals on $(a_0, a_1]$ and $[a_n, a_{n+1})$ may cause problems).

- Let $g : \mathbb{R} \to \mathbb{R}$ be a measurable function (i.e., $g^{-1}(B) \in \mathcal{R}$ for every $B \in \mathcal{R}$). Then, $E[g(X)] = \int_{\mathbb{R}} g(x) dF(x)$, provided the integral $\int_{\mathbb{R}} |g(x)| dF(x)$ exists (i.e., well-defined and finite).

  If X is continuous $E[g(X)] = \int_{\mathbb{R}} g(x) f(x) dx$, provided the integral $\int_{\mathbb{R}} |g(x)| f(x) dx$ exists.

A Word on Riemann Integrals $\int_a^b f(x)dx$



- Integral is the area under f.
- Partition $[a, b]$ by $P = \{x_0, x_1, \ldots, x_n\}$ with

$$a = x_0 < x_1 < \cdots < x_n = b$$

- Approximate integral by lower and upper Darboux sums:

$$L(f,P) = \sum_{i=0}^{n-1} \inf_{x\in[x_i,x_{i+1}]} f(x)(x_{i+1}-x_i)$$

$$U(f,P) = \sum_{i=0}^{n-1} \sup_{x\in[x_i,x_{i+1}]} f(x)(x_{i+1}-x_i)$$

- Approximation improves with finer partitions $\Rightarrow$ lower and upper Darboux integrals:

$$L(f) = \sup\{L(f,P) : P \text{ is a partition of } [a,b]\}.$$

$$U(f) = \inf\{U(f,P) : P \text{ is a partition of } [a,b]\}.$$

- Riemann integral is defined if $L(f) = U(f)$, and $\int_a^b f(x)dx = L(f) = U(f)$.

- Integral exists if f is continuous on $[a,b]$ or if it has countably many discontinuities.

A Word on Riemann-Stieltjes Integrals $\int_a^b f(x)dg(x)$

Suppose g is monotone and right-continuous.

Lower and upper sums:

$$\underline{S}(f,g;P) = \sum_{i=0}^{n-1} \inf_{x\in[x_i,x_{i+1}]} f(x)(g(x_{i+1}) - g(x_i)).$$

$$\bar{S}(f,g;P) = \sum_{i=0}^{n-1} \sup_{x\in[x_i,x_{i+1}]} f(x)(g(x_{i+1}) - g(x_i)).$$

Lower and upper integrals:

$$\bar{S}(f;g) = \inf\{\bar{S}(f,g;P) : P \text{ is a partition of } [a,b]\}.$$

$$\underline{S}(f,g) = \sup\{\underline{S}(f,g;P) : P \text{ is a partition of } [a,b]\}.$$

Riemann-Stieltjes integral is defined if $\underline{S}(f,g) = \bar{S}(f,g)$, and

$\int_a^b f(x)dg(x) = \underline{S}(f,g) = \bar{S}(f,g)$.

When $g(x) = x$, Riemann-Stieltjes integral becomes Riemann integral.

Riemann-Stieltjes integral generalizes in two ways:

- Generalized interval 'length' or 'measure': $g(x_{i+1}) - g(x_i)$
- Allows discontinuities in $g(x)$. For a monotone function, there are up to countably many.

  A discontinuity at t contributes $f(t)(g(t) - g(t_-))$ to the integral; whereas in the Riemann integral $f(t)(t - t_-) = 0$.

Existence: Both f and g can have countably many discontinuities, provided their discontinuities don't coincide.

To define Riemann-Stieltjes integral more generally, g need not be monotone. The usual requirement is that g has bounded variation. It is known that a function has bounded variation if and only if it can be written as the difference of two monotone functions. If, in addition, f is continuous on $[a, b]$, then the integral exists.

For a random variable X with the distribution function $F(x)$, $E[h(X)] = \int h(x)dF(x)$ covers the discrete, continuous and mixed random variables.

Riemann-Stieltjes integral is usually good enough for probability calculations, but not good enough for conceptualization.

Consider the definition of $E[X]$, $E[X] = \int_\Omega X(\omega)dP(\omega)$. What does it mean?

Ω is not necessarily $\mathbb{R}^k$. We need to define integral on abstract spaces. This leads to the general, measure-theoretic integral, sometimes called the Lebesgue integral.

Also, this general integration theory has rich structures, suggesting it is the right kind of generalization.

More to come.

Nonnegative Random Variables

Suppose X is a nonnegative random variable. Then,

$$E[X] = \int_0^\infty P(X \geq x)dx. \qquad \left(\text{Or } E[X] = \int_0^\infty P(X > x)dx\right).$$

Consider the indicator function $1_{\{X \geq x\}}$. Recall $E[1_{\{X \geq x\}}] = P(X \geq x)$.

$$\begin{aligned}
\int_0^\infty P(X \geq x)dx &= \int_0^\infty E[1_{\{X \geq x\}}]dx \\
&= \int_0^\infty \int_\Omega 1_{\{X \geq x\}}(\omega)dP(\omega)dx \\
&= \int_\Omega \int_0^\infty 1_{\{X \geq x\}}(\omega)dxdP(\omega) \\
&= \int_\Omega \int_0^{X(\omega)} 1 \, dxdP(\omega) \\
&= \int_\Omega X(\omega)dP(\omega) = E[X].
\end{aligned}$$

The interchange of the two integrals is justified by Fubini-Tonelli's Theorem (check it out). Here, the integrand $1_{\{X \geq x\}}(\omega)$ as a function of x and ω is a non-negative measurable function. That is enough.

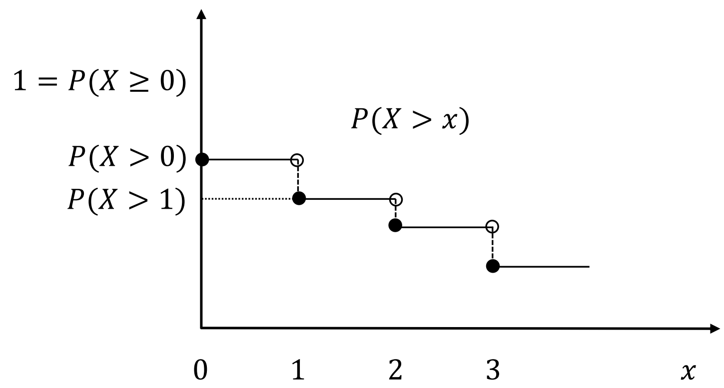


Figure 1: $\int_0^\infty P(X > x)dx$ is the area underneath the function $P(X > x)$

Suppose X is a nonnegative, integer-valued random variable. Then,

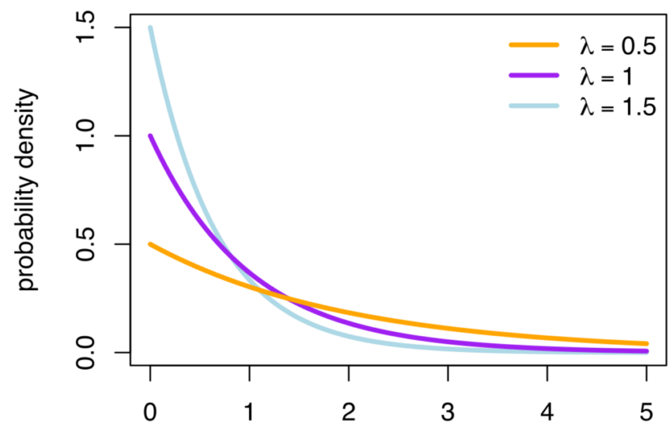
$$E[X] = \sum_{n=0}^{\infty} P(X > n) = \sum_{n=1}^{\infty} P(X \geq n).$$

In general, if X is a nonnegative random variable, then,

$$E[X] = \int_0^\infty \bar{F}(x)dx,$$

where $\bar{F}(x) = 1 - F(x)$ is the complementary cdf of X.

Exponential Distribution



Exponential distribution with parameter $\lambda > 0$: For $x \geq 0$,

$$F(x) = 1 - \exp(-\lambda x); \qquad f(x) = \lambda \exp(-\lambda x).$$

$E[X] = \int_0^\infty \bar{F}(x)dx = \int_0^\infty \exp(-\lambda x)dx = -\frac{1}{\lambda}\exp(-\lambda x)|_0^\infty = 1/\lambda$;
$\mathrm{var}(X) = 1/\lambda^2$.

An exponential random variable is often used to model the time between two consecutive events, e.g., arrival of buses, failures of equipment.

Memoryless Property: Suppose the inter-arrival time of buses has the exponential distribution with mean 10 minutes (i.e., $\lambda = 0.1$). Suppose you have waited for 9 minutes. What is the expected waiting time till the next bus arrival? Answer: 10 minutes.

Let X be the waiting time till the next bus arrival. For $x \geq 9$,

$$\begin{aligned} P(X > x | X > 9) &= \frac{P(X > x, X > 9)}{P(X > 9)} \\ &= \frac{\exp(-0.1x)}{\exp(-0.1 \times 9)} = \exp(-0.1(x-9)). \end{aligned}$$

Thus, when you shift the time origin to 9, i.e., let $t = x - 9$, the above conditional probability is equal to $\exp(-0.1t)$ for $t \geq 0$, which is exactly the complementary cdf of the exponential random variable with mean 10.

That is,

$$P(X > t + 9|X > 9) = P(X - 9 > t|X > 9) = \exp(-0.1t).$$

$X - 9$ is the additional or remaining waiting time till the next bus arrival.

Given you have waited for 9 minutes, the additional or remaining time till the next bus arrival is still exponentially distributed with the same mean.

This isn't a surprise as the exponential distribution is the continuous version (limiting case) of the geometric distribution.

Suppose a sequence of IID Bernoulli trials is performed at time $\delta, 2\delta, \ldots$. Let p be the success probability of each trial. Let W be the waiting time till the first success. Then,

$P(W > k\delta) = (1-p)^k$ for $k \geq 0$, or $P(W \geq k\delta) = (1-p)^{k-1}$ for $k \geq 1$.

Note $E[W] = \delta \sum_{k=0}^{\infty}(1-p)^k = \delta \sum_{k=1}^{\infty}(1-p)^{k-1} = \delta/p$.

For a fixed time t, $k = t/\delta$. Let $k \to \infty$ ($\delta \to 0$) and $p \to 0$ such that $p/\delta \to \lambda$ (hence, $p = \lambda\delta + o(\delta)$; also note $p/\delta = \frac{kp}{t}$, so the scaling of k and p is such that $kp \approx \lambda t$). Then,

$$P(W > t) = (1-p)^{t/\delta} \approx (1-\lambda\delta)^{t/\delta} \to \exp(-\lambda t).$$

In the above, we used $(1 - 1/n)^n \to e^{-1}$ with $1/n = \lambda\delta$.

Given the IID nature of the Bernoulli trials, it is not surprising that after waiting for some time without seeing a success, the remaining time till the first success has the same distribution as if we start from the first trial.

Gaussian/Normal Distribution

The pdf for Gaussian distribution with mean μ and variance σ^2:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), x \in \mathbb{R}.$$

The distribution function:

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

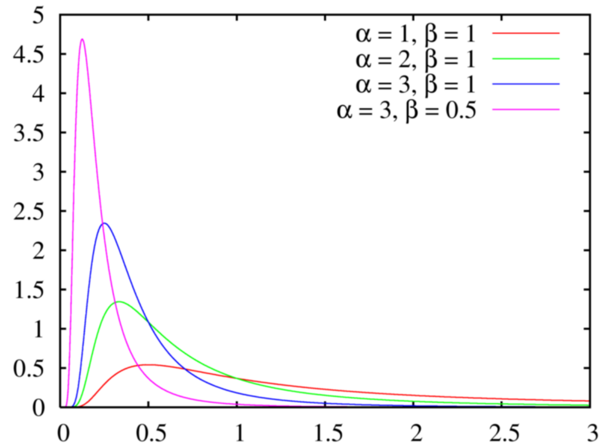
- It is denoted by $\mathcal{N}(\mu, \sigma^2)$.
- $\mathcal{N}(0,1)$ is called the standard normal distribution.

  Let X be $\mathcal{N}(\mu, \sigma^2)$. Then, $Y = (X-\mu)/\sigma$ is $\mathcal{N}(0,1)$.

  This can be verified by calculating $P(Y \le y) = P(X \le y\sigma + \mu)$.

  In general, given the above X, any $Y = \alpha X + \beta$ is $\mathcal{N}(\alpha\mu + \beta, \alpha^2\sigma^2)$ (prove this).

Gamma Distribution



- Gamma distribution (α, β): $\alpha, \beta > 0$; very useful for modeling

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \ \ x \geq 0.$$

The gamma function $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$. For integer n, $\Gamma(n) = (n-1)!$.

$E[X] = \frac{\alpha}{\beta}$, $\text{var}(X) = \frac{\alpha}{\beta^2}$.

- When α is an integer, we often rename $\alpha \leftrightarrow n, \beta \leftrightarrow \lambda$. Then,

$$f(x) = \frac{\lambda e^{-\lambda x}}{(n-1)!}(\lambda x)^{n-1}, \;\; x \geq 0.$$

In this form, it is also known as the Erlang distribution.

When $n = 1$, it becomes the exponential distribution.

The sum of n IID exponential random variables has the Erlang distribution:

If $X_1, \ldots, X_n$ are IID exponential random variables with parameter λ, then $X = X_1 + \ldots + X_n$ has the Erlang distribution (n, λ).

Other Examples

- Uniform on $[a, b]$: The density function $f(x) = 1/(b - a)$ on $[a, b]$ and zero elsewhere. $E[X] = (a + b)/2$.
- Pareto (α, K), where $\alpha, K > 0$:

$$F(x) = 1 - \left(\frac{K}{x}\right)^{\alpha}, \; f(x) = \frac{\alpha K^{\alpha}}{x^{\alpha+1}}, \;\; x \geq K.$$

$E[X] = \infty$ for $\alpha \leq 1$; $E[X] = \alpha K/(\alpha - 1)$ for $\alpha > 1$.
$\mathrm{var}(X) = \infty$ for $\alpha \leq 2$; $\mathrm{var}(X) = \frac{K^2 \alpha}{(\alpha-1)^2(\alpha-2)}$ for $\alpha > 2$.

It is the continuous version of the power law distribution.
Heavy-tail distribution: The tail of $\bar{F}(x)$ decays as a power function as apposed to exponential.
If you draw samples from such a heavy-tail distribution, large values

show up with a non-trivial probability.

If $(X_1, X_2, \ldots, X_n)$ are IID Pareto random variables, the maximum $(\max_i X_i)$ often dominates in the sum $(\sum_i X_i)$.

If you use Pareto distribution in simulation, you need to watch out the large values drawn from the Pareto distribution, especially when α is close (and above) 2. Sometimes the large values are so large that they don't make practical sense; furthermore, with those exceedingly large values, your simulation may not reach the steady-state behavior even after a long run. In those situations, you may need to use 'bounded' Pareto where the large values cannot exceed an upper bound. One possibility is that, whenever a value drawn from the original Pareto distribution exceeds the upper bound, you 'round' it down to the upper bound.

Joint Distribution

X and Y are defined on a common probability space $(\Omega, \mathcal{F}, P)$.

- The **joint distribution function** $F : \mathbb{R}^2 \to [0,1]$ of X and Y is

$$F(x,y) = P(X \leq x, Y \leq y).$$

Definition: X and Y are (jointly) continuous with joint probability density function $f : \mathbb{R}^2 \to [0,\infty)$ if

$$F(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u,v)dudv, \text{ for each } x, y \in \mathbb{R}.$$

If F is sufficiently differentiable at (x,y), we normally specify

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y).$$

Think of $f(x,y)dxdy$ as the (infinitesimal) probability of

$P(x < X \leq x + dx, y < Y \leq y + dy)$.

If $B \in \mathcal{R}^2$ is a sufficiently nice set, then

$$P((X,Y) \in B) = \int\int_B f(x,y)dxdy.$$

- The **marginal** distribution functions of X and Y are:

$$F_X(x) = P(X \leq x) = F(x,\infty); F_Y(y) = P(Y \leq y) = F(\infty, y).$$

The marginal density functions (when the joint density exists) are:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy; \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx.$$

- **Expectation of Continuous Random Variables:** Let $g : \mathbb{R}^2 \to \mathbb{R}$ be a measurable function (i.e., $g^{-1}(B) \in \mathcal{R}^2$ for every $B \in \mathcal{R}$). Then, $E[g(X,Y)] = \int_{\mathbb{R}}\int_{\mathbb{R}} g(x,y)f(x,y)dxdy$, provided the integral $\int_{\mathbb{R}}\int_{\mathbb{R}} |g(x,y)|f(x,y)dxdy$ exists (i.e., well-defined and finite).

- X and Y are independent if and only if

$$F(x, y) = F_X(x)F_Y(y), \text{ for all } x, y \in \mathbb{R},$$

and for the continuous case, if and only if

$$f(x, y) = f_X(x)f_Y(y), \text{ for all } x, y \in \mathbb{R}.$$

Conditional Distribution and Expectation

- For a discrete random variable, the conditional distribution function of Y given $X = x$ is defined by

$$F_{Y|X}(y|x) = P(Y \le y|X = x) = \frac{P(Y \le y, X = x)}{P(X = x)},$$

for any x such that $P(X = x) > 0$. This doesn't work for a continuous random variable because $P(X = x) = 0$ for all x. But, if the marginal density $f_X(x) > 0$ for some x, we have

$$P(Y \le y|x \le X \le x + dx) = \frac{P(Y \le y, x \le X \le x + dx)}{P(x \le X \le x + dx)}$$

$$\approx \frac{\int_{-\infty}^{y} \int_{x}^{x+dx} f(u, v)dudv}{f_X(x)dx} \approx \frac{\int_{-\infty}^{y} f(x, v)dxdv}{f_X(x)dx} = \int_{-\infty}^{y} \frac{f(x, v)}{f_X(x)} dv.$$

Definition: Suppose X and Y have the joint pdf $f(x, y)$. The

conditional distribution of Y given $X = x$, written as $F_{Y|X}(y|x)$ or $P(Y \leq y|X = x)$, is defined to be

$$F_{Y|X}(y|x) = \int_{-\infty}^{y} \frac{f(x, v)}{f_X(x)} dv,$$

for any x such that $f_X(x) > 0$. In such a case, $f_{Y|X}(y|x) \triangleq \frac{f(x,y)}{f_X(x)}$ is the **conditional density function**.

- When the conditional density exists, the **conditional expectation** of Y given $X = x$, written as $E[Y|X = x]$, is defined by $E[Y|X = x] = \int_{\mathbb{R}} y f_{Y|X}(y|x) dy$, for any x such that $f_X(x) > 0$. More generally,

$$E[Y|X = x] = \int_{\mathbb{R}} y dF_{Y|X}(y|x).$$

- $E[Y|X = x]$ is depends on x. Let $\psi(x) = E[Y|X = x]$. Then, $\psi(X) = E[Y|X]$ is a random variable, which depends X.

- **Important Fact:** We have $E[\psi(X)] = E[Y]$. In other words,

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

This works for all types of random variables.

Since $E[E[Y|X]] = E[\psi(X)]$, $E[E[Y|X]]$ means

$$E[E[Y|X]] = \int_{\mathbb{R}} \psi(x) dF_X(x) = \int_{\mathbb{R}} \int_{\mathbb{R}} y dF_{Y|X}(y|x) dF_X(x).$$

To compute $E[Y]$, we can conditional on some random variable first, say X, and compute the conditional expectation of Y given each $X = x$. Then, we 'average' the result over X (that is, compute the expectation of the result over the distribution of X).

For simplicity, suppose the conditional density and marginal density functions exist. Then,

$$\begin{aligned}\int_{\mathbb{R}}\int_{\mathbb{R}} y dF_{Y|X}(y|x) dF_X(x) &= \int_{\mathbb{R}}\int_{\mathbb{R}} y f_{Y|X}(y|x) f_X(x) dy dx \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} y f_{X,Y}(x,y) dy dx \\ &= \int_{\mathbb{R}} y \int_{\mathbb{R}} f_{X,Y}(x,y) dx dy \\ &= \int_{\mathbb{R}} y f_Y(y) dy \\ &= E[Y].\end{aligned}$$

Provided $E|Y| < \infty$, the interchange of integrals is justified by Fubini-Tonelli's theorem.

We can compute probabilities using the same 'conditioning' trick. Consider an event $B \in \mathcal{F}$. Let $Y = 1_B$, which is discrete. Suppose

X is a continuous random variable. Then,

$$\begin{aligned}
P(B) =& E[1_B] = E[E[1_B|X]] \\
=& \int_{\mathbb{R}} E[1_B|X = x] f_X(x) dx \\
=& \int_{\mathbb{R}} 1 \cdot P(1_B = 1|X = x) f_X(x) dx \\
=& \int_{\mathbb{R}} P(B|X = x) f_X(x) dx.
\end{aligned}$$

In the last step, we used $\{\omega \in \Omega : 1_B(\omega) = 1\} = B$. To go from step 2 to step 3, we are a bit fuzzy since we haven't defined what it means by $P(Y = y_i|X = x)$ when Y is discrete and X is continuous. Basically, it is the probability assignment for a discrete random variable, i.e., $\sum_i P(Y = y_i|X = x) = 1$. Then, we use the usual formula for calculating the expectation of a discrete random variable (i.e., $E[Z] = \sum_i z_i P(Z = z_i)$).

Sum of Random Variables

Suppose X and Y have joint density function f and $Z = X + Y$. Then,

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z - x)dx.$$

Formal method: Let $A = \{(x, y) : x + y \leq z\}$.

$$\begin{aligned}
P(Z \leq z) =& P(X + Y \leq z) = \int\int_A f(x, y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x, y)dydx, \quad \text{(change of variable } y = v - x\text{)} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{z} f(x, v - x)dvdx \\
&= \int_{-\infty}^{z}\int_{-\infty}^{\infty} f(x, v - x)dxdv.
\end{aligned}$$

Therefore, $\int_{-\infty}^{\infty} f(x, v - x)dx$ is the density function for Z.

Heuristic argument:

$$\begin{aligned} f_Z(z)dz \approx& P(z \leq X + Y \leq z + dz) \\ =& \int_{-\infty}^{\infty} \int_{z-x}^{z-x+dz} f(x, y)dydx \\ \approx& \int_{-\infty}^{\infty} f(x, z - x)dzdx. \end{aligned}$$

Canceling dz, we get the desired $f_Z(z)$.

Next, suppose X and Y are independent. Then,
$f(x, z - x) = f_X(x)f_Y(z - x)$, and

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z - x)dx = \int_{-\infty}^{\infty} f_Y(y)f_X(z - y)dy.$$

f_Z is equal to the convolution of f_X and f_Y, written as $f_Z = f_X * f_Y$.

Example

Suppose X_1 and X_2 are independent exponential random variables with parameter λ. Suppose $X = X_1$ and $S = X_1 + X_2$. What is the conditional distribution of X given $S = s$?

First, we need to compute the joint density of X and S, $f(x, s)$. The joint distribution function of X and S is $F(x, s) = P(X_1 \le x, X_1 + X_2 \le s)$.

Note that when $x > s$,

$$\begin{aligned} F(x, s) =& P(X_1 \le s, X_1 + X_2 \le s) + P(s < X_1 \le x, X_1 + X_2 \le s) \\ =& P(X_1 \le s, X_1 + X_2 \le s). \end{aligned}$$

Therefore, we will assume $x \le s$ in the following.

$$
\begin{aligned}
F(x,s) =& P(X_1 \le x, X_1 + X_2 \le s) \\
=& \int_0^x P(X_1 + X_2 \le s | X_1 = u) f_{X_1}(u) du \\
=& \int_0^x P(X_2 \le s - u) f_{X_1}(u) du \\
=& \int_0^x (1 - e^{-\lambda(s-u)}) \lambda e^{-\lambda u} du \\
=& 1 - e^{-\lambda x} - \lambda x e^{-\lambda s}
\end{aligned}
$$

We used the trick of 'conditioning' (on X_1) to get the second equality.

$$
f(x,s) = \frac{\partial^2 F}{\partial x \partial s}(x,s) = \lambda^2 e^{-\lambda s}, \;\; 0 \le x \le s.
$$

The marginal density of S is

$$
f_S(s) = \int_0^\infty f(x,s) dx = \int_0^s f(x,s) dx = \lambda^2 s e^{-\lambda s}.
$$

Then, for $0 \leq x \leq s$,

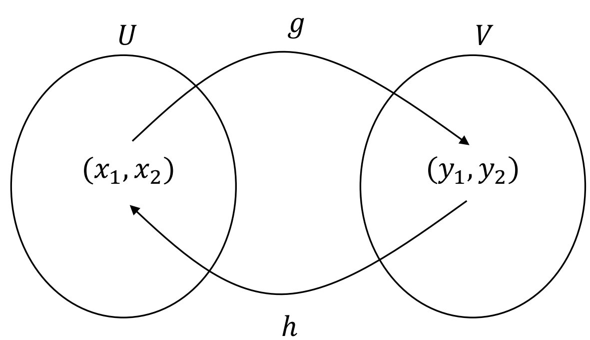
$$P(X \leq x|S = s) = \int_0^x f_{X|S}(u|s)du$$
$$= \int_0^x \frac{f(u, s)}{f_S(s)}du$$
$$= \frac{x}{s}.$$

Conclusion: Given $X_1 + X_2 = s$, X_1 is uniformly distributed on $[0, s]$.

Functions of Random Variables

Suppose X_1 and X_2 have joint density function f. Suppose $g_1, g_2 : \mathbb{R}^2 \to \mathbb{R}$ are measurable functions. Let $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ What is the joint density function of Y_1 and Y_2?

Let the mapping $(x_1, x_2) \mapsto (y_1, y_2) = (g_1(x_1, x_2), g_2(x_1, x_2))$ be denoted by $g = (g_1, g_2)$. For simplicity, suppose g is a one-to-one and onto mapping from U to V, where $U, V \subseteq \mathbb{R}^2$. Suppose $(X_1, X_2) \in U$ and $(Y_1, Y_2) \in V$.



Then, $(x_1, x_2) = h(y_1, y_2)$, where $h = g^{-1}$. Suppose h can be written as

$h(y_1, y_2) = (h_1(y_1, y_2), h_2(y_1, y_2))$. The Jacobian matrix of h is

$$J_h = \begin{pmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} \\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2} \end{pmatrix}$$

For a typical set $B \subseteq V$, let $A = g^{-1}(B) = h(B) \in U$. Then,

$$\begin{aligned} P((Y_1, Y_2) \in B) =& P((X_1, X_2) \in A) = \int\int_A f(x_1, x_2) dx_1 dx_2 \\ &= \int\int_B f(h(y_1, y_2)) \left|\text{det}(J_h(y_1, y_2))\right| dy_1 dy_2. \end{aligned}$$

In the above, we used the change of variable formula in calculus, and $\text{det}(J_h)$ is the determinant of the Jacobian matrix.

We see that the joint density function of (Y_1, Y_2) is

$$f_{Y_1,Y_2}(y_1, y_2) = f(h_1(y_1, y_2), h_2(y_1, y_2)) \left|\text{det}(J_h(y_1, y_2))\right|,$$

for (y_1, y_2) in the range of g.

Rewrite the earlier example as $Y_1 = X_1$ and $Y_2 = X_1 + X_2$. We have $g(x_1, x_2) = (x_1, x_1 + x_2)$ and g maps the set $\{(x_1, x_2) : x_1 \geq 0, x_2 \geq 0\}$ onto the set $\{(y_1, y_2) : 0 \leq y_1 \leq y_2\}$.

Then, $X_1 = Y_1$ and $X_2 = Y_2 - Y_1$. The inverse function $h = (h_1, h_2)$ is given by $h_1(y_1, y_2) = y_1$ and $h_2(y_1, y_2) = y_2 - y_1$. Then,

$$J_h = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix},$$

and $\det(J_h(y_1, y_2)) = 1$. Then,

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) =& f(y_1, y_2 - y_1) = \lambda e^{-\lambda y_1} \lambda e^{-\lambda(y_2 - y_1)} \\ =& \lambda^2 e^{-\lambda y_2}, \text{ for } 0 \leq y_1 \leq y_2. \end{aligned}$$