Week 7

Probability Refresher

7.1 Random Variables

Recall that r.v. X is a function defined on a Probability Space $(\Omega, \mathcal{F}, \mathbb{P})$,

$$X : \Omega \to \mathbb{R}$$

with the property that for any $x \in \mathbb{R}$, $\{\omega : X(\omega) \leqslant x\} \in \mathcal{F}$, i.e., for any $x \in \mathbb{R}$, we can calculate the probability

$$\mathbb{P}(\{\omega : X(\omega) \leqslant x\}) \overset{\text{denote for short}}{=} \mathbb{P}(X \leqslant x),$$

i.e., we will be able to define one of the most important characteristics of our r.v. X, the Cumulative Distribution Function

$$F(x) = F_X(x) = \mathbb{P}(X \leqslant x), \qquad x \in \mathbb{R}.$$

7.1.1 Discrete Random Variables

Definition 7.1. *We say that our r.v. X is **discrete**, if the range of X, i.e., the set of all values of X, is finite or countably infinite.*

Note: Please note that Ω itself can be uncountable, but X can be discrete. Say, we can have $\Omega = [0, 1]$ and $X(\omega) = 2$ for any $\omega \in \Omega$, so the range of X consists of only one point, $\{2\}$. ■

For a discrete r.v. X, we define

- The CDF $F(x)$ by
$$F(x) = F_X(x) = \mathbb{P}(X \leqslant x), \qquad x \in \mathbb{R};$$
- The PMF $f(x)$ by[1]
$$f(x) = f_X(x) = \mathbb{P}(X = x), \qquad x \in \mathbb{R}.$$

Now, having the PMF of the discrete r.v. X with possible values $x_1, x_2, ...$, we can calculate the probability $\mathbb{P}(X \in A)$ for any subset $A \subset \mathbb{R}$:

$$\mathbb{P}(X \in A) = \sum_{x_i \in A} \mathbb{P}(X = x_i).$$

The most important discrete random variables are:

[1] That is, if the range of X (the possible values of X) is $\{x_1, x_2, ...\}$, then

$$f(x_i) = \mathbb{P}(X = x_i), \qquad i = 1, 2, ...$$

One sometimes defines also $f(x) = 0$ for all $x \notin \{x_1, x_2, ...\}$.

Discrete Uniform Distribution

Parameters: $n \in \mathbb{N}$, range $x_1, x_2, ..., x_n \in \mathbb{R}$

Notation: $X \sim \mathsf{DiscreteUnif}(n; x_1, ..., x_n)$

Range: $\{x_1, ..., x_n\}$

PMF:

Value of X	x_1	x_2	...	x_n
PMF of X	$\frac{1}{n}$	$\frac{1}{n}$	...	$\frac{1}{n}$

Expectation and Variance: $\mathbb{E}(X) = \overline{x} = \frac{x_1+x_2+...+x_n}{n}$, $Var(X) = \frac{\sum_{k=1}^{n} x_k^2}{n} - \left(\frac{\sum_{k=1}^{n} x_k}{n}\right)^2$

Usage: Describes Equiprobable Phenomena, say, dice rolling

PMF Plot:

R code for PMF Plot:

Additional Properties:

Bernoulli Distribution

Parameters: $p \in [0, 1]$

Notation: $X \sim \mathsf{Bernoulli}(p)$

Range: $\{0, 1\}$

PMF:

Values of X	0	1
$\mathbb{P}(X = x)$	$1 - p$	p

One can write this in the form (we will use this form later, say, in the Maximum Likelihood Method)

$$f(x) = p^x \cdot (1-p)^{1-x}, \qquad x \in \{0, 1\}.$$

Expectation and Variance: $\mathbb{E}(X) = p$, $Var(X) = p(1-p)$

Usage: Describes Phenomena with two outcomes, say, "success" and "failure", or "yes" and "no", or "boy" and "girl", "smaller or equal" and "larger"; 1 designates the success (or "yes" or ...), and the probability of success is p

PMF Plot:

R code for PMF Plot:

Additional Properties:

Binomial Distribution

Parameters: $n \in \mathbb{N}$, $p \in [0, 1]$

Notation: $X \sim \mathrm{Binomial}(n, p)$ or $X \sim \mathrm{Binom}(n, p)$ or $X \sim \mathcal{B}(n, p)$

Range: $\{0, 1, 2, ..., n\}$

PMF:

Values of X	0	1	2	...	n
$\mathbb{P}(X = x)$	$\binom{n}{0} p^0 (1-p)^{n-0}$	$\binom{n}{1} p^1 (1-p)^{n-1}$	$\binom{n}{2} p^2 (1-p)^{n-2}$	...	$\binom{n}{n} p^n (1-p)^{n-n}$

Expectation and Variance: $\mathbb{E}(X) = n \cdot p$, $Var(X) = n \cdot p(1-p)$

Usage: Describes Phenomena when we do a certain experiment several times, we are dealing with independent trials of the same experiment. Here n is the number of trials[2], p is the probability of success in each of the trials. Here X shows the number of successes in n trials.

PMF Plot:

R code for PMF Plot:

Additional Properties:

- Assume $X_1, ..., X_m$ are independent and $X_k \sim \mathrm{Bernoulli}(p)$. Then
$$X_1 + X_2 + ... + X_m \sim \mathrm{Binomial}(m, p).$$
- Assume $X_k \sim \mathrm{Binomial}(n_k, p)$ for $k = 1, ..., m$ and X_k are independent. Then
$$\sum_{k=1}^{m} X_k \sim \mathrm{Binomial}(n_1 + ... + n_m, p).$$

Geometric Distribution

Parameters: $p \in (0, 1)$

Notation: $X \sim \mathrm{Geom}(p)$

Range: $\mathbb{N}$

PMF:

Value of X	1	2	3	...	n	...
PMF of X	p	$p(1-p)$	$p(1-p)^2$	...	$p(1-p)^{n-1}$	...

Expectation and Variance: $\mathbb{E}(X) = \frac{1}{p}$, $Var(X) = \frac{1-p}{p^2}$

Usage: Describes Phenomena when we do a certain experiment several times, we are dealing with independent trials of the same experiment. p is the probability of success in each of the trials. Here X shows the number of flips needed until the first success[3].

Example: Assume an AUA student knows what is the Geometric Distribution with probability p. We stand at the AUA entrance and ask all entering students who knows what is the Geometric Distribution. Let X be the random variable showing the number if the first student who will answer "Yes, I know what is the GD". Then $X \sim \mathrm{Geom}(p)$.

Example: Assume that the probability that some part of a car will fail in a particular day is p. What is the probability that the part will last at least n days?

[2] If you are rolling 10 times a die, then $n = 10$.

[3] There is another version of Geometric Distribution in the literature, the number of **failures** Y before the first success. It can be seen that $Y = X - 1$, so it is easy to find the PMF of Y.

PMF Plot:

R code for PMF Plot:

Additional Properties:

- Assume $X_1, ..., X_m, ...$ are independent and $X_k \sim \mathrm{Bernoulli}(p)$. Then

$$X = \min\{k \in \mathbb{N} : X_k = 1\} \sim \mathrm{Geom}(p).$$

- (Memoryless Property) If $X \sim \mathrm{Geom}(p)$, then for every non-negative integers x, t,

$$\mathbb{P}(X > x + t | X > t) = \mathbb{P}(X > x).$$

Poisson Distribution

Parameters: $\lambda > 0$

Notation: $X \sim \mathrm{Poisson}(\lambda)$ or $X \sim \mathrm{Pois}(\lambda)$

Range: $\mathbb{N} \cup \{0\}$

PMF:

Value of X	0	1	2	...	n	...
PMF of X	$e^{-\lambda} \cdot \frac{\lambda^0}{0!}$	$e^{-\lambda} \cdot \frac{\lambda^1}{1!}$	$e^{-\lambda} \cdot \frac{\lambda^2}{2!}$	...	$e^{-\lambda} \cdot \frac{\lambda^n}{n!}$	...

Expectation and Variance: $\mathbb{E}(X) = \lambda$, $Var(X) = \lambda$

Usage: Describes Phenomena when we do repeatidely some experiment, and want to calculate the number of success within some bounded interval, and we know that we are dealing with a rare event (in the sense that n is large and np is small[4]). For example, we want to model the number of calls in a call center within a 10-min interval. Or we want to count the number of page click in Google page for some webpage within a day.

Example: The number of visits to some fixed website today can be modelled as a Poisson r.v. . The parameter here is the mean number of visits during today (usually, one can take for the parameter the average of daily visits during the several previous days).

Example: The number of claims per day for some insurance company. The parameter here is the average number of claims per day.

PMF Plot:

R code for PMF Plot:

Additional Properties:

- λ is the average number of events in an interval of interest
- Poisson Distribution is the limit of Binomial Distribution where $n \to \infty$ and $p \to 0$ such that

$$\lambda = p \cdot n = \mathrm{const}.$$

- Assume $X_k \sim \mathrm{Poisson}(\lambda)$ for any $k = 1, ..., n$, and X_k are independent. Then

$$X_1 + ... + X_n \sim \mathrm{Poisson}(n \cdot \lambda).$$

[4] Say, $n \geqslant 50$ and $np \leqslant 5$

7.1.2 Continuous Random Variables

Definition 7.2. *We say that our r.v.* X *is* ***continuous*** *or absolutely continuous, if there exists an integrable, non-negative function* $f(x)$, $x \in \mathbb{R}$ *such that*

$$F(x) = \int_{-\infty}^{x} f(t)dt, \qquad x \in \mathbb{R}.$$

Here, $f(x)$ stands for the PDF of our r.v.

Now, if X is a continuous r.v. with PDF $f(x)$, then for (almost) any $A \subset \mathbb{R}$,

$$\mathbb{P}(X \in A) = \int_{A} f(x)\,dx,$$

say, in particular,

$$\mathbb{P}(a \leqslant X \leqslant b) = \int_{a}^{b} f(x)\,dx, \qquad \text{for any} \quad [a, b],$$

and

$$\mathbb{P}(X = a) = 0, \qquad \text{for any} \quad a \in \mathbb{R}.$$

It can be proved that the real function $f : \mathbb{R} \to \mathbb{R}$ is a PDF of some continuous r.v. if and only if

$$f(x) \geqslant 0 \quad \text{for almost all } x \in \mathbb{R} \qquad \text{and} \qquad \int_{-\infty}^{\infty} f(x)\,dx = 1.$$

7.1.3 Uniform Distribution

Parameters: a, b with $a < b$

Notation: $X \sim Unif(a, b)$ or $X \sim \mathcal{U}(a, b)$

PDF and CDF:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b] \\ 0, & \text{otherwise} \end{cases} \qquad F(x) = \begin{cases} 0, & \text{if } x < a \\ \dfrac{x-a}{b-a}, & \text{if } x \in [a, b] \\ 1, & \text{if } x > b \end{cases}$$

Expectation and Variance: $\mathbb{E}(X) = \frac{a+b}{2}$, $Var(X) = \frac{(b-a)^2}{12}$

Usage: Describes Phenomena like "picking a point at random from an interval", where the choice of picking points are "equiprobable". In fact, the probability of picking each point is 0, so one needs to describe this as follows. The probability that we are picking from $[a, b]$ is 1. The probability that our point will be on the left or right halves is 0.5. We divide our interval into 4 equal-length parts. The probability to have our point in each of this intervals is 0.25 etc. So, in general, the probability to have our point in a given subinterval is proportional to the length of that subinterval (but does not depend on the position of that interval).

Example: Meeting a regular metro train.

If $a = 0$, $b = 1$, then we deal with Standard Uniform Distribution.

PDF and CDF Plots:

R code for PDF and CDF Plots:

Additional Properties:

7.1.4 Exponential Distribution

Parameters: $\lambda > 0$

Notation: $X \sim \text{Exp}(\lambda)$

PDF and CDF: $f(x) = \begin{cases} \lambda \cdot e^{-\lambda x}, & \text{if } x \geqslant 0 \\ 0, & \text{if } x < 0 \end{cases} \qquad F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{if } x \geqslant 0 \\ 0, & \text{if } x < 0 \end{cases}$

Expectation and Variance: $\mathbb{E}(X) = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$

Usage: Describes Phenomena like modelling the time elapsed until the occurrence of certain event, or the time between events (waiting times), when that time is random.

λ- the rate parameter is the average "arrival rate", the reciprocal of the average time between the events,

$$\lambda = \frac{1}{\text{average time between events}}.$$

Relation to the Poisson Distribution: If we are calculating the lengths of the inter-arrival times for a Poisson Distributed r.v., then that lengths are Exponentially Distributed.

Is some sense, Exponential Distribution is the continuous counterpart of the Geometric Distribution, the limiting case, when the number of trials is very large, and the probability of success is very small.

Example: The amount of time between two consecutive visits for some particular website can be modelled as an Exp r.v.

Example: The time between two consecutive earthquakes can be modelled as an Exp r.v.

Example: The arrival of students to the Stat lecture can be modelled as a Poisson r.v., and the interarrival times are Exponentially Distributed.

Example: The lifetime of some component of something

PDF and CDF Plots:

R code for PDF and CDF Plots:
```
par(mfrow = c(1,2))
curve(pexp(x, rate = 1), xlim = c(-1,3), lwd = 3, col = "blue")
curve(dexp(x, rate = 1), xlim = c(-1,3), lwd = 3, col = "blue")
```

Additional Properties:

- (Memoryless Property) If $X \sim \text{Exp}(\lambda)$, and if $x, t \geqslant 0$, then
$$\mathbb{P}(X \geqslant x + t | X \geqslant t) = \mathbb{P}(X \geqslant x).$$
- If $X \sim \text{Exp}(\lambda)$ and $\alpha \in \mathbb{R}$, $\alpha \neq 0$, then $\alpha \cdot X \sim \text{Exp}(\frac{\lambda}{\alpha})$
- See a lot of more at https://en.wikipedia.org/wiki/Exponential_distribution

7.1.5 Normal (Gaussian) Distribution

Parameters: μ - mean, σ^2 - variance

Notation: $X \sim \mathcal{N}(\mu, \sigma^2)$

PDF and CDF: $f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad F(x) = \int_{-\infty}^{x} f(t)dt$

Expectation and Variance: $\mathbb{E}(X) = \mu$, $Var(X) = \sigma^2$

Usage: Describes a lot of phenomena. Important is the CLT.

If $\mu = 0$ and $\sigma = 1$, then $X \sim \mathcal{N}(0,1)$ is called a Standard Normal Random Variable, and usually denoted by Z. In that case also one uses the notations $\phi(x)$ and $\Phi(x)$ for the Standard Normal PDF and CDF, respectively.

PDF and CDF Plots:

R code for PDF and CDF Plots:

Additional Properties:

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Z = \dfrac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$ and

$$\mathbb{P}(a < X < b) = \mathbb{P}\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

- If $Z \sim \mathcal{N}(0,1)$, then $X = \mu + \sigma \cdot Z \sim \mathcal{N}(\mu, \sigma^2)$
- If $X_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$ for $k = 1, ..., n$ are independent, then

$$X_1 + X_2 + ... + X_n \sim \mathcal{N}(\mu_1 + \mu_2 + ... + \mu_n, \sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2).$$

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\mathbb{P}(-\sigma < X - \mu < \sigma) \approx 0.6827, \qquad \mathbb{P}(-2\sigma < X - \mu < 2\sigma) \approx 0.9545$$

and

$$\mathbb{P}(-3\sigma < X - \mu < 3\sigma) \approx 0.9973.$$

Concerning the Normal Distribution, there is another nice property: it maximizes the Entropy. To define the entropy, let us start by giving a measure to estimate the level of Surprise or Uncertainty.

Assume in the result of some experiment some event E can happen. We want to give some numerical measure to the how surprised we will be if E occurs. Of course, that "Surprise Measure" need to relate the probability of event E. So if p is the probability of E to happen, then the surprise is a function of p, let us denote it by[5] $S(p)$, $p \in (0,1]$. Now, to construct S, we impose some intuitive properties, that we want to be satisfied by S:

S1: $S(1) = 0$, i.e., if $p = 1$, then we are not surprised that E happened;

S2: S strictly decreases on $(0,1]$, i.e., if $p_1 < p_2$, then $S(p_1) > S(p_2)$. This means that if we have two events E_1 and E_2 with probabilities p_1 and p_2, respectively, then if $p_1 < p_2$, the surprise that E_2 happened is less than of E_1 happened (since E_1 is more unlikely to happen);

S3: S is continuous, i.e., small change in p results to a small change in surprise measure;

S4: $S(p \cdot q) = S(p) + S(q)$, for all $p, q \in (0,1]$, i.e., if E_1 and E_2 are two independent events with probabilities $\mathbb{P}(E_1) = p$ and $\mathbb{P}(E_2) = q$, then $\mathbb{P}(E_1 \cap E_2) = p \cdot q$. Now, $S(p \cdot q)$ is the surprise measure that $E_1 \cap E_2$ happened. Since E_1 and E_2 are independent, this property states that the surprise that E_1 and E_2 happened is just the sum of individual surprises of that of E_1 happened and E_2 happened. That is,

$$\text{Surprise from } E_1 \text{ and } E_2 \overset{E_1 \text{ and } E_2 \text{ are indep}}{=\!=\!=\!=} \text{Surprise from } E_1 + \text{Surprise from } E_2.$$

[5] We will not define $S(p)$ for $p = 0$ at this moment.

Now, using these 4 properties, one can prove that

$$S(p) = -C \cdot \log_2(p),$$

for some constant C.

Usually, one takes $C = 1$, and in that case we say that $S(p) = -\log_2(p)$ is measuring the surprise in bits.

Now, if we have a discrete r.v. X, which can take the values $x_1, x_2, ..., x_n$ with probabilities $p_1, p_2, ..., p_n$, correspondingly, then we define[6]

$$H(X) = -\sum_{k=1}^{n} p_k \cdot \log_2(p_k),$$

and we call $H(X)$ the entropy of r.v. X, the measure of expected amount of surprise we will receive when getting the value of X. In information theory, $H(X)$ is interpreted as the average amount of information received when the value of X is observed, or the amount of uncertainty that exists as to the value of X, the amount how "unpredictable" is X. The measure units for $H(X)$ are **bits**.

Example: Assume X us a random variable that can take 16 different equiprobable values. Then to specify the outcome, the value of X, one needs 4 bits.

Now, let us calculate the entropy of this r.v.:

$$H(X) = -\sum_{k=1}^{16} \frac{1}{16} \cdot \log_2\left(\frac{1}{16}\right) = \log_2 16 = 4.$$

So the entropy coincides with the (average) number of bits necessary to represent the outcome of X. ■

Example: Assume we are tossing a coin, and we know that the probability of tails is p, and for heads is $q = 1 - p$. Then the Entropy of the Bernoulli r.v. for this experiment will be maximized if $p = 0.5$, the maximum unpredictability, the maximum uncertainty is when $p = 0.5$. This intuitive fact can be proven by maximizing

$$H(X) = -(p \log_2 p + (1-p) \log_2(1-p)), \qquad p \in [0, 1]. \quad ■$$

Example: See example 2.1.2 From Thomas Cover, Joy Thomas, *"Elements of Information Theory"*, p. 15.

Now, if X is a continuous r.v. with PDF $f(x)$, then we define the entropy as[7]

$$H(X) = -\int_{-\infty}^{+\infty} f(x) \cdot \ln f(x)\, dx,$$

where we assume $0 \cdot \ln 0 = 0$.

Now, going back to our Normal Distribution, if we will consider all continuous r.v. X with given $\mathbb{E}(X) = \mu$ and $Var(X) = \sigma^2$, then the one that maximizes the entropy is the Normal Distribution, i.e., among all r.v. X with $\mathbb{E}(X) = \mu$ and $Var(X) = \sigma^2$, the entropy $H(X)$ is maximal iff $X \sim \mathcal{N}(\mu, \sigma^2)$.

[6] In Greek, H is the capital E=Eta letter.

[7] One can use also the base-2 logarithms, and this will be a constant multiple of this one given above.

7.1.6 Gamma Distribution

We define first one of the important mathematical special functions, Euler's Gamma function:

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} \cdot e^{-x} dx, \qquad \alpha \in (0, +\infty).$$

It can be shown, that Γ function has the following properties:

Proposition 7.1. *(Properties of the Γ function)*

a. $\Gamma(\alpha+1) = \alpha \cdot \Gamma(\alpha)$ *for any* $\alpha > 0$*;*

b. $\Gamma(n+1) = n!$*, for any* $n = 0, 1, 2, ...$*, so* Γ *function is the extension of the factorial notion for all positive numbers*[8]

c. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

d. $\Gamma(x) \cdot \Gamma(1-x) = \dfrac{\pi}{\sin \pi x}$ *for any* $x \in (0, 1)$

Now we define the $\mathtt{Gamma}$-distribution:
Notation: $X \sim \mathrm{Gamma}(\alpha, \beta)$ or $X \sim \Gamma(\alpha, \beta)$
PDF:

$$f(x|\alpha, \beta) = \begin{cases} \dfrac{1}{\beta^{\alpha} \cdot \Gamma(\alpha)} \cdot x^{\alpha-1} e^{-x/\beta}, & x > 0 \\ 0, & x \leqslant 0 \end{cases}$$

Some properties:
If $X \sim \mathrm{Gamma}(\alpha, \beta)$, then $\mathbb{E}(X) = \alpha \cdot \beta$ and $\mathrm{Var}(X) = \alpha \cdot \beta^2$;
If $X_k \sim \mathrm{Gamma}(\alpha_k, \beta)$ are independent, then $X_1 + ... + X_n \sim \mathrm{Gamma}(\alpha_1 + ... + \alpha_n, \beta)$;
If $X \sim \mathrm{Gamma}(1, \frac{1}{\lambda})$, then $X \sim \mathrm{Exp}(\lambda)$
If $X \sim \mathrm{Gamma}(n/2, 2)$, then $X \sim \chi^2(n)$
Applications: (from Wiki, `https://en.wikipedia.org/wiki/Gamma_distribution`)
The gamma distribution has been used to model the size of insurance claims and rainfalls.

The gamma distribution is also used to model errors in multi-level Poisson regression models, because the combination of the Poisson distribution and a gamma distribution is a negative binomial distribution.

In wireless communication, the gamma distribution is used to model the multi-path fading of signal power.

In neuroscience, the gamma distribution is often used to describe the distribution of inter-spike intervals.

In bacterial gene expression, the copy number of a constitutively expressed protein often follows the gamma distribution, where the scale and shape parameter are, respectively, the mean number of bursts per cell cycle and the mean number of protein molecules produced by a single mRNA during its lifetime.

In genomics, the gamma distribution was applied in peak calling step (i.e. in recognition of signal) in ChIP-chip and ChIP-seq data analysis.

The gamma distribution is widely used as a conjugate prior in Bayesian statistics. It is the conjugate prior for the precision (i.e. inverse of the variance) of a normal distribution. It is also the conjugate prior for the exponential distribution.

[8] This means, that if we will see something like $(3.2)!$, we can understand this as $\Gamma(4.2)$.

7.1.7 Beta Distribution

Here we first introduce the Euler's Beta function $\mathcal{B}$:

$$\mathcal{B}(\alpha, \beta) = \int_0^1 x^{\alpha-1} \cdot (1-x)^{\beta-1} dx, \qquad \alpha, \beta \in (0, +\infty).$$

The important relation between Euler's Gamma and Beta functions is

$$\mathcal{B}(\alpha, \beta) = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha+\beta)} \qquad \forall \alpha, \beta > 0.$$

Now, about the Beta distribution:
Notation: $X \sim \mathrm{Beta}(\alpha, \beta)$ or $X \sim \mathcal{B}(\alpha, \beta)$
PDF:

$$f(x|\alpha, \beta) = \begin{cases} \dfrac{1}{\mathcal{B}(\alpha, \beta)} \cdot x^{\alpha-1}(1-x)^{\beta-1}, & x \in (0,1) \\ 0, & x \notin (0,1) \end{cases}$$

Note that if $X \sim \mathcal{B}(\alpha, \beta)$, then X can take only values in $[0,1]$ in the sense that $\mathbb{P}(X \notin [0,1]) = 0$.
Properties:
If $X \sim \mathrm{Beta}(\alpha, \beta)$, then $\mathbb{E}(X) = \frac{\alpha}{\alpha+\beta}$ and $\mathrm{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

7.2 Some Operations with R.V.s

7.2.1 Transformation of r.v.'s

Assume now that we have a r.v. X and we have a function $g : \mathbb{R} \to \mathbb{R}$, and we define a new r.v. Y, the transformation of X by g, by

$$Y = g(X).$$

Our aim is to express the CDF F_Y and PMF/PDF f_Y of Y in terms of ones of X. The discrete case is simple[9], so we will talk only about the continuous case.

Unfortunately, for general case of g, there is no formula to express F_Y and f_Y in terms of F_X and f_X, respectively. But we have the following result:

Theorem 7.1. *If X is a r.v. with CDF F_X and PDF f_X, and if $g : \mathbb{R} \to \mathbb{R}$ is strictly monotone differentiable function, then for the r.v. $Y = g(X)$, we will have:*

- *If g is strictly increasing, then $F_Y(y) = F_X(g^{-1}(y))$, for any $y \in \mathbb{R}$;*
- *If g is strictly decreasing, then $F_Y(y) = 1 - F_X(g^{-1}(y))$, for any $y \in \mathbb{R}$.*

Also, in both cases,

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left|(g^{-1}(y))'\right|, \qquad y \in \mathbb{R}.$$

Now, assume we want to calculate the expected value of the r.v. $g(X)$. We have 2 methods to do this: one of them is to use the PDF of $g(X)$:

$$\mathbb{E}(g(X)) \overset{Y=g(X)}{=} \mathbb{E}(Y) = \int_{-\infty}^{+\infty} y \cdot f_Y(y) dy = \int_{-\infty}^{+\infty} x \cdot f_{g(X)}(x) dx.$$

To calculate by this formula, we need to calculate first the PDF $f_{g(X)}$. Fortunately, we have an other way, a great formula, to use only the PDF of X:

$$\mathbb{E}(g(X)) = \int_{-\infty}^{+\infty} g(x) \cdot f_X(x) dx.$$

Nice and smooth and clear and great!

[9] Do that by yourself!

7.3 Joint Distribution of Random Variables

Example: Assume X is the highest price for Google (Alphabet Inc.) stock and Y is the S&P500 index highest value for tomorrow. The X and Y are r.v.'s defined on the same probability space (which is the set of all possible scenarios of the market for tomorrow).

If we have two r.v.'s X and Y defined on the same probability space, then we can talk about their Joint CDF:

$$F(x,y) = F_{X,Y}(x,y) = \mathbb{P}(X \leqslant x, Y \leqslant y), \qquad x,y \in \mathbb{R}.$$

Recall that we have defined the independence of r.v.'s X and Y in the following way:

Definition 7.3. *We say that r.v.'s X and Y, defined on the same probability space, are independent, if*

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B), \qquad \textit{for any subsets } A, B \subset \mathbb{R}.$$

It can be proved that this is equivalent to

$$F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y) \qquad \forall x, y \in \mathbb{R}.$$

Proposition 7.2. *If X and Y are independent r.v., then for any functions h and g, the r.v. $h(X)$ and $g(Y)$ are independent.*

7.3.1 Discrete Case

For discrete case of X and Y, we talk also about their Joint PMF:

$$f(x,y) = f_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y), \qquad x \in Range(X), \quad y \in Range(Y).$$

This can be represented in the form of a table: assume the range of X is $x_1, x_2, ...$ and the range of Y is $y_1, y_2, ...$ (along this paragraph, we will assume this are the possible values of X and Y, respectively). Then the Joint PMF table is:

!! Give here the table, with marginals. The sum of all entries in the main part =1

Having this table, we can calculate probabilities like $\mathbb{P}((X,Y) \in A)$ for any $A \subset \mathbb{R}^2$:

$$\mathbb{P}((X,Y) \in A) = \sum_{(x_i, y_j) \in A} \mathbb{P}(X = x_i, Y = y_j), \qquad \forall A \subset \mathbb{R}^2.$$

Now, in the case of discrete r.v.'s X and Y, if the range of X is $x_1, x_2, ...$, and the range of Y is $y_1, y_2, ...$, they will be independent if and only if

$$\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i) \cdot \mathbb{P}(Y = y_j), \qquad \text{for all } i \text{ and } j,$$

so our Joint PMF Table for X and Y is:

!! Write the table with $\mathbb{P}(X = x_i) \cdot \mathbb{P}(Y = y_j)$, and the marginals $\mathbb{P}(X = x_j)$, $\mathbb{P}(Y = y_j)$.

Now assume our discrete r.v.'s X and Y have the above Joint PMF, and we form $Z = X + Y$. Then Z will be discrete, of course. We want to express the PMF of Z in terms of ones of X and Y.

Proposition 7.3. *If X and Y are independent discrete r.v. with ranges $x_1, x_2, ...$ and $y_1, y_2, ...$, respectively, then the range of Z will be a subset of*[10] $\{x_k + y_j\}$*, and for z from that range,*

$$\mathbb{P}(Z = z) = \sum_i \mathbb{P}(X = x_i) \cdot \mathbb{P}(Y = z - x_i).$$

[10] Maybe not the whole set $x_i + y_j$. Say, it can happen that $X(\omega_1) = x_1$ and $Y(\omega_1) = y_1$, so $Z(\omega_1) = x_1 + y_1$, and if $X(\omega) \neq x_1$ for any $\omega \neq \omega_1$, then Z will not take the value $x_1 + y_2$.

Probability Distribution of $g(X,Y)$

Assume now we have a function $g : \mathbb{R}^2 \to \mathbb{R}$, and we form the r.v. $g(X,Y)$. Our aim here is to find the probability distribution of that r.v..

Clearly,

Values of $g(X,Y)$	$g(x_1,y_1)$	$g(x_1,y_2)$	$g(x_2,y_1)$	$g(x_1,y_3)$	...	$g(x_i,$
PMF of $g(X,Y)$	$\mathbb{P}(X=x_1,Y=y_1)$	$\mathbb{P}(X=x_1,Y=y_2)$	$\mathbb{P}(X=x_2,Y=y_1)$	$\mathbb{P}(X=x_1,Y=y_3)$	...	$\mathbb{P}(X=x_i,$

Here we assume that if for different i and j, the values of $g(x_i,y_j)$ coincide, then we add the corresponding probabilities.

Example:

Calculation of Expectations

If X and Y are discrete r.v.'s with ranges $x_1, x_2, ...$ and $y_1, .y_2, ...$, and if $g : \mathbb{R}^2 \to \mathbb{R}$ is a function, then

$$\mathbb{E}(g(X,Y)) = \sum_{x_i,y_j} g(x_i,y_j) \cdot \mathbb{P}(X = x_i, Y = y_j).$$

7.3.2 Continuous Case

In the case of continuous X and Y their joint behaviour is again described by their Joint CDF:

$$F(x,y) = F_{X,Y}(x,y) = \mathbb{P}(X \leqslant x, Y \leqslant y), \qquad x,y \in \mathbb{R}.$$

This is the same as for the Discrete case. Now, we will call the pair (the vector) (X,Y) continuous, if they have a Joint PDF $f(x,y)$, i.e., an integrable function satisfying

$$\mathbb{P}((X,Y) \in A) = \iint_A f(x,y)dxdy, \qquad A \subset \mathbb{R}^2.$$

In particular,

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)dudv.$$

Having the joint distribution of X and Y, one can invoke the individual distributions of X and Y:

$$F_X(x) = F_{X,Y}(x,+\infty), \quad \forall x \in \mathbb{R}, \qquad \text{and} \qquad F_Y(x) = F_{X,Y}(+\infty,y), \quad \forall y \in \mathbb{R}.$$

Also, if the pair (X,Y) is continuous, then so are the r.v.'s X and Y, and their PDF's, called Marginal PDF's for X and Y, will be

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy, \quad \forall x \in \mathbb{R},$$

and

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx, \quad \forall y \in \mathbb{R}.$$

Note: Please note that the inverse is note true! If X and Y are separately continuous, then not necessarily (X,Y) will be continuous.

!! Here give an example!

Independence

Again, we have defined the independence of r.v.'s X and Y in the following way:

Definition 7.4. *We say that r.v.'s X and Y, defined on the same probability space, are independent, if*

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A) \cdot \mathbb{P}(Y \in B), \qquad \textit{for any subsets} A, B \subset \mathbb{R}.$$

In the case of continuous vector (X, Y), we will have:

Proposition 7.4. *The following assertions are equivalent:*

a. X *and* Y *are independent;*

b. $F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y)$ *for any* $x, y \in \mathbb{R}$;

c. $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$ *for any* $x, y \in \mathbb{R}$,

where F_X, F_Y, f_X *and* f_Y *are marginal CDF and PDF for* X *and* Y, *respectively.*

Note: As we have stated above, having the Joint Distribution of X and Y, we can find the individual distributions of X and Y. But if we know the individual distributions of X and Y, it is **not possible**, in general, to find the distribution of (X, Y). In the particular case when X and Y are independent, we can find the Joint Distribution by the above Proposition. In the general case it is not possible. The problem is that besides the individual distributions of X and Y, one needs to know also the relationship between X and Y.

Proposition 7.5. *If* X *and* Y *are independent, then*

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

and

$$Var(X + Y) = Var(X) + Var(Y).$$

Distribution of the sum of two r.v.'s

Now assume that (X, Y) is continuous random vector, and we form $Z = X + Y$. Then it can be proven that Z will be a continuous r.v.. We want to express the PDF of Z in terms of ones of X and Y. The next Theorem is giving the formulae:

Theorem 7.2. *If* (X, Y) *is a continuous r.vector with Joint PDF* $f(x, y)$, *then the PDF* f_{X+Y} *of r.v.* $X + Y$ *will be*

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f(x, z - x)dx, \qquad \forall z \in \mathbb{R}.$$

In the particular case, if X *and* Y *are independent continuous r.v. with PDFs* f_X *and* f_Y, *then the PDF of* $Z = X + Y$ *will be*

$$f_{X+Y}(z) = \int\limits_{-\infty}^{+\infty} f_X(z - v)f_Y(v)dv = \int\limits_{-\infty}^{+\infty} f_X(v)f_Y(z - v)dv, \qquad \forall z \in \mathbb{R}.$$

Calculation of Expectations

If (X, Y) is continuous random vector with Joint PDF $f(x, y)$, and if $g : \mathbb{R}^2 \to \mathbb{R}$, then[11]

$$\mathbb{E}(g(X, Y)) = \iint_{\mathbb{R}^2} g(x, y) \cdot f(x, y) dxdy.$$

!Give here examples, calc of covariance, exp of product etc.

Example: By considering $g(x, y) = x$ or $g(x, y) = y$, one will get

$$\mathbb{E}(X) = \iint_{\mathbb{R}^2} x \cdot f(x, y) dxdy \qquad \text{and} \qquad \mathbb{E}(Y) = \iint_{\mathbb{R}^2} y \cdot f(x, y) dxdy.$$

7.4 Some Additions

Question 1: Assume $X, Y \sim \mathtt{Bernoulli}(0.5)$, and they are independent. What is the distribution of $X + Y$?

Answer: $X + Y \sim \mathtt{Binom}(2, 0.5)$.
This can be seen by using the PMF.

Question 2: Assume $X, Y \sim \mathtt{Bernoulli}(0.5)$. What can be said about the distribution of $X + Y$, is it Binomial?

Answer: No, in general. For example, if $Y = X$, then $X + X = 2X$, and $2X$ assumes the values 0 and 2, both with probability 0.5, and this is not a binomial distribution.

Question 3: Assume X is a r.v., $a, b \in \mathbb{R}$. Is it true that $Y = a \cdot X + b$ have the same distribution as X?

- If $X \sim \mathtt{Binom}(n, p)$, is it true that $a \cdot X + b \sim \mathtt{Binom}(m, p')$ for some m, p' ?
- If $X \sim \mathtt{Exp}(\lambda)$, is it true that $X + b \sim \mathtt{Exp}(\lambda')$ for some λ' ?
- If $X \sim \mathtt{Exp}(\lambda)$, is it true that $a \cdot X \sim \mathtt{Exp}(\lambda')$ for some λ' ?
- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then it can be proved that $a \cdot X + b$ will be Normally distributed. Find with which parameters.
- Which families of distributions are invariant wrt linear transformations?

Question 4: Assume X and Y are normal r.v. (with, say, the same parameters). Is it true that $X + Y$ is again normal (with some parameters)?

Question 5: Let $X_{(i)}$ be the i-th order statistics of $X_1, ..., X_n$, where X_k-s are IID r.v.. Find the distribution of $X_{(i)}$. In particular, find the distributions of

$$X_{(1)} = \min\{X_1, ..., X_n\} \qquad \text{and} \qquad X_{(n)} = \max\{X_1, ..., X_n\}.$$

[11] If, of course, the integral converges!