Module 2

Random variables

Topics: Random variables, Distribution, mass, and density functions, Common random variables, Functions of a random variable, Expected value, Variance and quantiles

Random variables

Consider two random experiments - rolling a fair die and randomly selecting two balls from a pool of two white and two black balls, denoted by W_1, W_2 and B_1, B_2 respectively. Note that $\Omega_1 = \{1,2,3,4,5,6\}$ and $\Omega_2 = \{W_1W_2, W_1B_1, W_1B_2, W_2B_1, W_2B_2, B_1B_2\}$. We assumed that the balls are selected simultaneously in the 2nd experiment. Classical measure applies to both the experiments. With this we can calculate probability and conditional probability involving any events and check independence of events. In addition to these, in the first experiment, we can calculate the (long-run) mean outcome of the roll, i.e., $\lim_{n\to\infty} \sum_{i=1}^{n} r_i/n$ where r_i denotes outcome of the i-th roll. Later, we will prove that this mean outcome is same as the weighted sum of the outcomes with the weights being probabilities of the outcomes. Then the mean outcome is: $\sum_{i=1}^{6} i \cdot P(\{i\}) = \sum_{i=1}^{6} i \cdot (1/6) = 3.5$. Later, we will formally define this mean outcome as expected value, along with concepts like variance, quantiles, etc.

In the second experiment, we may be interested in obtaining the (long-run) mean number of white balls selected. However, the concept of mean outcome is undefined for Ω_2, as it does not contain numbers. We can overcome this difficulty by considering a function X that maps every outcome of Ω to the number of white balls corresponding to the outcome, i.e., $X: \Omega_2 \to \mathbb{R}$ such that $X(W_1W_2) = 2$, $X(W_1B_1) = X(W_1B_2) = X(W_2B_1) = X(W_2B_2) = 1$, $X(B_1B_2) = 0$. Observe that Ω_2 has been transformed into $X(\Omega_2) = \{2,1,0\}$, which can be considered as a random experiment. It is appropriate to measure probability for $X(\Omega_2)$ using that of Ω_2 as: $P_{X(\Omega_2)}(A) = P_{\Omega_2}(\{\omega \in \Omega_2 : X(\omega) \in A\})$ for all $A \subseteq X(\Omega_2)$, i.e., probability of $A \in X(\Omega_2)$ is same as the probability of those outcomes in Ω_2 that are mapped to an element of A by X. Let us calculate probability of $A = \{1\}$ in $X(\Omega_2)$. $W_1B_1, W_1B_2, W_2B_1, W_2B_2 \in \Omega_2$ are mapped to $\{1\}$ by X. Then $P_{X(\Omega_2)}(\{1\}) = P_{\Omega_2}(\{W_1B_1, W_1B_2, W_2B_1, W_2B_2\}) = 4/6$. Similarly, one can obtain $P_{X(\Omega_2)}(\{0\}) = P_{X(\Omega_2)}(\{2\}) = 1/6$. Then the (long-run) mean number of white balls selected can be calculated as: $\sum_{i=0}^{2} i \cdot P_{X(\Omega_2)}(\{i\}) = 1 \times (4/6) + 2 \times (1/6) = 1$. *Consider selection of n balls from a pool of w white and b black balls. Assuming $n \leq b + w$, obtain the probability of $k \leq \min(w, n)$ white balls selected.*

We formalize the transformation of Ω_2 into $X(\Omega_2)$ through the concept of random variable, which is defined as a real-valued function on the sample space of a random experiment, i.e., $X: \Omega \to \mathbb{R}$. We define probability measure for the new sample space $X(\Omega)$ using that of Ω. For $A \subseteq X(\Omega)$, we define $P_{X(\Omega)}(A) = P_\Omega(\{\omega \in \Omega : X(\omega) \in A\})$. If no transformation of Ω is needed, e.g., the first experiment, then X can be considered as the identity function. *Consider the experiment of breaking a 1m long stick at an arbitrary point. Let Y denote the length of the longer piece. Obtain $P(\{Y \leq y\})$ for $y \in (0.5,1)$.* Random variable allows us to define concepts like expected value, variance, quantiles, etc. and it permits the use of calculus and other tools for manipulating numbers, which is not possible with all sample spaces. Random variable enables us to extract more meaning out of a random experiment.

Distribution, mass, and density functions

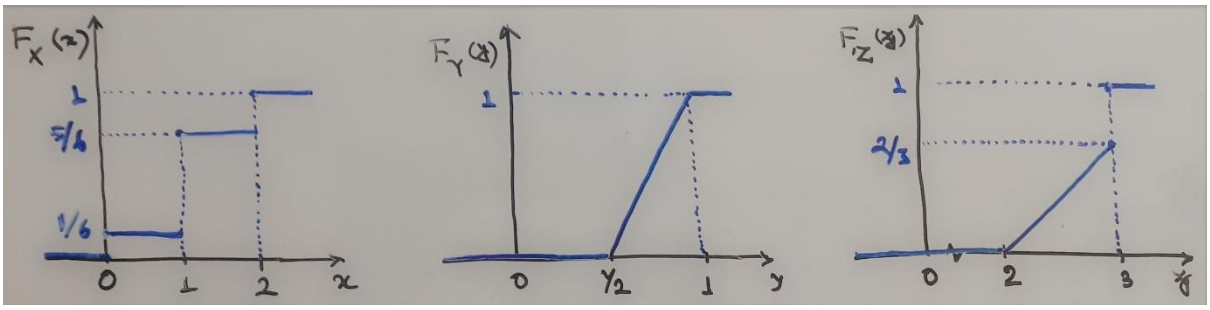
We define (cumulative) distribution function for a random variable as: $F: \mathbb{R} \to [0,1]$ such that $F(x) = P_{X(\Omega)}\big((-\infty, x]\big) = P_{\Omega}(\{\omega \in \Omega: X(\omega) \le x\})$. Distribution function is a fundamental feature of a random variable. We can use it to calculate probability of any event associated with a random variable. Since $X(\Omega) \subseteq \mathbb{R}$, then any event $A \subseteq X(\Omega)$ can be represented as the union of disjoint open intervals and singletons. Let $A = (\cup_{i=1}^{m} I_i) \cup \big(\cup_{j=1}^{n} \{a_j\}\big)$, where I_i denotes an open interval. By the third axiom, $P_{X(\Omega)}(A) = \sum_{i=1}^{m} P_{X(\Omega)}(I_i) + \sum_{j=1}^{n} P_{X(\Omega)}\big(\{a_j\}\big)$. Using the distribution function, we can calculate these probabilities as follows:

$$(-\infty, a] = (-\infty, a) \cup \{a\} \text{ and these are disjoint} \Rightarrow \mathrm{P}_{X(\Omega)}(\{a\})$$
$$= \mathrm{P}_{X(\Omega)}\big((-\infty, a]\big) - \mathrm{P}_{X(\Omega)}\big((-\infty, a)\big) = \mathrm{F(a)} - \lim_{x \to a^-} F(x) = \mathrm{F(a)} - \mathrm{F}(a^-)$$

$$\text{Let } I = (b, c);\ (-\infty, c) = (-\infty, b] \cup (b, c) \text{ and these are disjoint} \Rightarrow \mathrm{P}_{X(\Omega)}\big((b, c)\big)$$
$$= \mathrm{P}_{X(\Omega)}\big((-\infty, c)\big) - \mathrm{P}_{X(\Omega)}\big((-\infty, b]\big) = \lim_{x \to c^-} F(x) - \mathrm{F(b)} = \mathrm{F}(c^-) - F(b)$$

where $\mathrm{F}(a^-)$ denotes the left-limit of $F(x)$ at $x = a$. *Using the above principle, obtain $P_{X(\Omega)}(I)$ in terms of distribution function when I is not an open interval.*

Let X denote the number of white balls when two balls are randomly selected from a pool of 2 white and 2 black balls. *Verify that* $F_X(x) = 0$ *for* $x < 0$, $1/6$ *for* $x \in [0,1)$, $5/6$ *for* $x \in [1,2)$, *and* 1 *for* $x \ge 2$. Let Y denote the length of the longer piece when a 1m long stick is broken into two pieces randomly. *Verify that* $F_Y(y) = 0$ *for* $y < 0.5$, $2y - 1$ *for* $y \in [0.5,1]$, *and* 1 *for* $y > 1$. Let Z denote the time taken by a student to write a 3-hour exam. Consider that the student would take any time between 2 and 3.5 hours to write the paper, with all values in $\Omega = [2,3.5]$ being equally likely, when there is no time limit. Note that classical measure applies to Ω. With the time limit, we can write $Z = \min(t, 3)$ where t is the outcome of the random experiment Ω. *Verify that* $F_Z(z) = 0$ *for* $z < 2$, $(z - 2)/1.5$ *for* $z \in [2,3)$, *and* 1 *for* $z \ge 3$. These three distribution functions are depicted below.



In all three cases, one can observe the following properties of the distribution function: (i) $F(-\infty) = 0$ and $F(+\infty) = 1$, (ii) $F(x)$ is non-decreasing in $x \in \mathbb{R}$, and (iii) $F(x)$ is right-continuous at all $x \in \mathbb{R}$. *If you are not familiar with the concepts like right-continuity or left-limit (mentioned earlier), do some readings*. These three properties are true for all random variables. We can prove the first two properties quite easily. Since X maps outcomes to real numbers, $X(\omega) \in (-\infty, \infty)\ \forall \omega \in \Omega$. Then $F(-\infty) = P_\Omega(\{\omega \in \Omega : X(\omega) \leq -\infty\}) = P_\Omega(\emptyset) = 0$ and $F(+\infty) = P_\Omega(\{\omega \in \Omega : X(\omega) \leq +\infty\}) = P_\Omega(\Omega) = 1$. For the second property, consider $x_1 < x_2$. Since $(-\infty, x_2] = (-\infty, x_1] \cup (x_1, x_2]$ and these are disjoint, $P_{X(\Omega)}\big((-\infty, x_2]\big) = P_{X(\Omega)}\big((-\infty, x_1]\big) + P_{X(\Omega)}\big((x_1, x_2]\big) \geq P_{X(\Omega)}\big((-\infty, x_1]\big) \Rightarrow F(x_1) \leq F(x_2)$. We can verify the third property for a given random variable, but a proof is beyond our scope.

If $X(\Omega)$ is countable, we refer to the random variable as discrete. A discrete random variable can be defined on any kind of sample space, whereas a non-discrete random variable, i.e., when $X(\Omega)$ is uncountable, can be defined only on uncountable sample spaces. We define (probability) mass function for a discrete random variable as: $p: X(\Omega) \to [0,1]$ such that $p(x) = P_{X(\Omega)}(\{x\}) = P_\Omega(\omega \in \Omega : X(\omega) = x)$. For discrete random variables, mass function is as fundamental as the distribution function, and it is easier to calculate than the distribution function. We can obtain one function from the other as follows: $F(a) = \sum_{x \in X(\Omega) \cap (-\infty, a]} p(x)$ for all $a \in \mathbb{R}$ and $p(x) = F(x) - F(x^-)$ for all $x \in X(\Omega)$. *Consider roll of two fair dice. Let X denote the sum of the outcomes. Obtain the mass function of X.*

Distribution function of a discrete random variable is discontinuous at some points, while it can be continuous throughout in $\mathbb{R}$ or discontinuous at some points for a non-discrete random variable. If distribution function of a random variable is continuous throughout in $\mathbb{R}$, we refer to it as continuous. A random variable that is neither discrete nor continuous is referred to as a mixed random variable. Such a random variable has uncountable $X(\Omega)$ and discontinuity of distribution function at some points. Every distribution function increases from 0 to 1 as we move from $-\infty$ to $+\infty$. For discrete random variables, this increase happens via jumps alone, for continuous random variables, this happens via continuous increase, and for mixed random variables, this happens partly via jumps and partly via continuous increase.

We define (probability) density function for continuous random variables, which is analogous to the mass function for the discrete case, as: $f: \mathbb{R} \to [0, \infty)$ such that $F(a) = \int_{-\infty}^{a} f(x)dx$ for all $a \in \mathbb{R}$. It can be verified that the derivative of $F(x)$, if exists throughout in $\mathbb{R}$, qualifies as a density function. Since $F(x)$ is defined on $\mathbb{R}$ and its derivative exists throughout in $\mathbb{R}$, then $F'(x)$ is defined on $\mathbb{R}$. Since $F(x)$ is non-decreasing in $x \in \mathbb{R}$, then $F'(x)$ is non-negative. Thus, $F': \mathbb{R} \to [0, \infty)$. The last requirement, i.e., $\int_{-\infty}^{a} F'(x)dx = F(a)\ \forall a \in \mathbb{R}$, is implied by the fundamental theorem of calculus. *If you are not familiar with this result, do some readings*. If $F(x)$ is not differentiable at countably many $x \in \mathbb{R}$, then a density function can be obtained as: $f(x) = F'(x)$ whenever $F'(x)$ exists, else $f(x) =$ any non-negative finite number. If $F(x)$ is not differentiable at uncountable number of points, then density function does not exist. The claims made in the last two sentences require measure theoretic results, which are beyond the scope of this course.

Consider the experiment of breaking a 1m long stick at an arbitrary point. Let Y denote the length of the longer piece. Earlier, we obtained that $F(y) = 0$ for $y < 0.5$, $2y - 1$ for $y \in [0.5,1]$, and 1 for $y > 1$. We can see that $F(y)$ is differentiable at all points in $\mathbb{R}$ except 0.5 and 1. $F'(y) = 0$ for $y < 0.5$, 2 for $y \in (0.5,1)$, and 0 for $y > 1$. Then $f(y) = 2$ for $y \in (0.5,1)$ and 0 everywhere else qualifies as a density function of Y. *Let r denote revenue of a firm and c denote its cost. Consider (r, c), in crore rupee, is equally likely to be any point in* $[1,2] \times [0.8,1.4]$. *Let X denote the profit. Obtain density function of X.*

We can obtain density function directly from Ω, though it may not be the convenient way. For sufficiently small $\delta > 0$, we can write the following:

$$P_{X(\Omega)}\big((a - \delta/2\,, a + \delta/2]\big) = P_{X(\Omega)}\big((-\infty, a + \delta/2]\big) - P_{X(\Omega)}\big((-\infty, a - \delta/2]\big)$$
$$= F(a + \delta/2) - F(a - \delta/2) = \int_{-\infty}^{a+\delta/2} f(x)dx - \int_{-\infty}^{a-\delta/2} f(x)dx$$
$$= \int_{a-\delta/2}^{a+\delta/2} f(x)dx \approx f(a)\delta$$

The above approximation becomes exact when $\delta \to 0^+$. Therefore,

$$f(a) = \lim_{\delta\to 0^+} \frac{P_{X(\Omega)}\big((a - \delta/2\,, a + \delta/2]\big)}{\delta} = \lim_{\delta\to 0^+} \frac{P_\Omega(\{\omega \in \Omega : X(\omega) \in (a - \delta/2\,, a + \delta/2]\})}{\delta}$$

Unlike mass functions, density function does not represent probability. It can take value more than 1, which mass function cannot. However, density can be used to represent probability that a continuous random variable X takes value in an infinitely small interval of length dx around x. From the above discussion, this probability is $f(x)dx$. Since $dx \to 0^+$, the interval around x can be thought of as the point x itself, and then $P(X = x) = \lim_{dx\to 0^+} f(x)dx = 0$, as $f(x)$ is finite. So, the probability of a continuous random variable taking a specific value is always zero, which is not the case with discrete random variables.

Common random variables

Now we talk about some commonly occurring random variables. We begin with the simplest one – the uniform random variable. Consider X to be the identity function defined on $\Omega \subset \mathbb{R}$, where classical measure applies. For example, consider rolling of a fair die and our interest is in the number obtained, or consider breaking of a 1m long stick at an arbitrary point and we are interested in one of the pieces. In the first case, $\Omega = \{1,2,3,4,5,6\}$, and in the second case, $\Omega = [0,1]$. In both the cases, classical measure applies to Ω, and $X(\omega) = \omega \; \forall \omega \in \Omega$. Note that the classical measure applies to $X(\Omega)$ as well. In the discrete case, we can write $X(\Omega) = \{v_1, v_2, \dots, v_n\}$, then the mass function is given by: $p(v_i) = 1/n$ for $i = 1,2, \dots, n$. In the continuous case, for example if $X(\Omega) = [a, b]$, then the distribution function is given by:

$F(x) = 0$ for $x < a$, $(x-a)/(b-a)$ for $x \in [a,b]$, and 1 for $x > b$. Its density function is given by: $f(x) = 1/(b-a)$ for $x \in [a,b]$, and 0 otherwise.

If we can classify outcomes of a random experiment into two categories – success and failure, we refer to the experiment as Bernoulli trial. We denote probability of success in a Bernoulli trial by p. A binomial random variable counts the number of successes in n independent and identical Bernoulli trials. We can express $\Omega = \{S,F\}_1 \times \{S,F\}_2 \times \cdots \times \{S,F\}_n$, and $X(\Omega) = \{0,1,2,\dots,n\}$. Consider x number of successes. Any x out of the n trials can lead to success, while the remaining $n-x$ trials lead to failure. So, there are $C(n,x)$ different ways of having x success in n trials, and each of these outcomes of Ω has probability $p^x(1-p)^{n-x}$. So, the mass function is given by: $p(x) = C(n,x)p^x(1-p)^{n-x}$ for $x = 0,1,\dots,n$.

A geometric random variable counts the number of independent and identical Bernoulli trials required to get the first success. Here, $\Omega = \{S, FS, FFS, \dots\}$ and $X(\Omega) = \{1,2,3,\dots\}$. Consider first success in x-th trial. Then the first $x-1$ trials must result in failure. So, the probability of first success in x-th trial is $(1-p)^{x-1}p$. So, the mass function is: $p(x) = (1-p)^{x-1}p$ for $x = 1,2,3,\dots$. Geometric random variable is memoryless, i.e., how further we must wait to get the first success is not influenced by how long we have already waited. Mathematically, we need $P(X > s+t | X > s) = P(X > t)$ for all $s,t > 0$ for a positive random variable X to be memoryless. For geometric a random variable, $P(X > x) = \sum_{y=x+1}^{\infty} p(y) = (1-p)^x$ for all $x = 1,2,3,\dots$ (*verify this*). Then for all $s,t \in \{1,2,3,\dots\}$,

$$P(X > s+t | X > s) = \frac{P(X > s+t, X > s)}{P(X > s)} = \frac{P(X > s+t)}{P(X > s)} = \frac{(1-p)^{s+t}}{(1-p)^s} = (1-p)^t$$
$$= P(X > t), \text{as required.}$$

Poisson process captures behavior of naturally occurring events over time or space. Consider the decay of a piece of radio-active material through α-particle radiation. Let $N(t)$ denote the number of α-particle emitted during $(0,t]$. Let us divide $(0,t]$ into n sub-intervals of equal length. Let X_i denote the number of α-particle emitted during the i-th sub-interval. Then $N(t) = X_1 + X_2 + \cdots + X_n$. *If two events cannot occur at the same point in time*, then by making n arbitrarily large, we can ensure that $P(X_i \geq 2)$ is negligible for all i. Let $\lambda_i = \lim_{n\to\infty} P(X_i = 1)/(t/n)$ denote the rate of occurrence of an event in the i-th sub-interval. *If this rate does not change with time*, i.e., $\lambda_i = \lambda$ for all i, then $X_1, X_2, \dots, X_n$ can be thought of as identical Bernoulli trials with success probability $P(X_i = 1) = \lambda t/n$. Finally, *if the sub-intervals behave independently*, then $N(t)$ follows the binomial distribution with parameters n and $\lambda t/n$, where $n \to \infty$. Let us take the limit. For $x \in \{0,1,2,\dots\}$,

$$P(N(t) = x) = \lim_{n\to\infty} C(n,x)\left(\frac{\lambda t}{n}\right)^x \left(1 - \frac{\lambda t}{n}\right)^{n-x} = \lim_{n\to\infty} \frac{n!}{x!\,(n-x)!} \frac{(\lambda t/n)^x}{(1-\lambda t/n)^x} \left(1 - \frac{\lambda t}{n}\right)^n$$
$$= \frac{(\lambda t)^x}{x!} \lim_{n\to\infty} \frac{n(n-1)\cdots(n-x+1)}{n^x} \frac{1}{(1-\lambda t/n)^x} e^{-\lambda t}$$
$$= e^{-\lambda t} \frac{(\lambda t)^x}{x!} \lim_{n\to\infty} \frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-x+1}{n} = e^{-\lambda t} \frac{(\lambda t)^x}{x!}.$$

Hence, the mass function of a Poisson random variable with rate λ and duration t is given by: $p(x) = e^{-\lambda t}\,(\lambda t)^x/x!$ for $x = 0,1,2, \ldots.$ Observe that the duration t can take any value. So, a Poisson random variable is characterized by its rate λ alone.

Exponential distribution describes the time between consecutive events when the events are occurring in a natural setting and Poisson process is applicable. Let T denote the time till the first event. Then for $t > 0$, events $\{T > t\}$ and $\{N(t) = 0\}$ are the same, and therefore, $P(T > t) = P(N(t) = 0) = e^{-\lambda t}$. So, the distribution function is given by: $F(t) = 0$ for $t \leq 0$ and $1 - e^{-\lambda t}$ for $t > 0$. Its density is given by $f(t) = \lambda e^{-\lambda t}$ for $t > 0$ and 0 otherwise. Due to the assumptions of steady rate of occurrence of events and independent behavior of sub-intervals, one can consider a fresh beginning every time an event takes place. Then the time between consecutive events is equivalent to the time till the first event, which follows exponential distribution with rate λ. Like the geometric distribution, exponential distribution too is memoryless (*verify this*). Imagine you are waiting in a bus stop, and the busses arrive in such a manner that Poisson process is applicable. Then the time you already waited in the bus stop has no bearing on how much longer you must wait.

Unlike other random variables, where we provided a context first, we are going to define the normal random variable through its density function. Later, we will see one context where it arises. The density function of a normal random variable with parameters μ and σ^2 is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \text{ for all } x \in \mathbb{R}, \text{where } \sigma^2 > 0$$

Since we did not describe Ω, $X(\Omega)$, etc., we do not know whether the above density function is a valid density or not. It is evident that $f(x) \geq 0$ for all $x \in \mathbb{R}$. Now, if we can prove that $\int_{-\infty}^{\infty} f(x)dx = 1$, then $f(x)$ is a valid density function. *When will you say that a given mass function or a distribution function is valid? Verify that the mass function of Poisson random variable and distribution function of exponential random variable are valid.*

$$\text{Let } I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-u^2} du \text{ by replacing } \frac{x-\mu}{\sqrt{2}\sigma} = u$$

$$\Rightarrow I^2 = \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-u^2} du\right) \times \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-v^2} dv\right) = \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(u^2+v^2)} dudv.$$

The double integration is over a plane, which is covered by rectangular grids of dimensions du and dv, where u and v both ranges from $-\infty$ to ∞. If we switch to polar coordinates, i.e., replace $u = r\cos\theta$ and $v = r\sin\theta$, then the plane can be covered through annulus grids of dimensions $rd\theta$ and dr, where θ ranges from 0 to 2π and r ranges from 0 to ∞. Then

$$I^2 = \frac{1}{\pi} \int_0^{2\pi} \int_0^{\infty} e^{-r^2} r dr d\theta = \frac{1}{\pi} \int_0^{2\pi} d\theta \int_0^{\infty} \frac{1}{2} e^{-s} ds \text{ by replacing } r^2 = s$$

$$\Rightarrow I^2 = \frac{1}{\pi} \times 2\pi \times \left(\frac{e^{-s}}{-2}\right)_0^{\infty} = 2\left(-\frac{e^{-\infty}}{2} + \frac{e^0}{2}\right) = 1 \Rightarrow I = 1, \text{as required.}$$

We do not have any closed-form expression for the distribution function of a normal random variable, as $F(x) = \int_{-\infty}^{x} f(y)dy$ does not simplify. Numerical evaluation is a way out. It can be done efficiently by converting a given normal random variable into the standard normal random variable, where $\mu = 0$ and $\sigma^2 = 1$. We discuss this in the next section.

Functions of a random variable

Let X denote demand of a product. It's uncertain and can be reasonably modelled by a normal random variable with parameters μ and σ^2. Then, in short, we write $X \sim N(\mu, \sigma^2)$. Let p, v, c denote unit selling price, unit production cost, and fined cost of production. Then, profit can be written as: $Y = (p - v)X - c$. Note that Y is a real-valued function defined on $X(\Omega)$, i.e., $Y: X(\Omega) \to \mathbb{R}$, where Ω is the original sample space and $X(\Omega) \subseteq \mathbb{R}$ is the X-transformation of Ω. Observe that Y fits the definition of a random variable, with $X(\Omega)$ as the original sample space and $Y\big(X(\Omega)\big) \subseteq \mathbb{R}$ is the Y-transformation of $X(\Omega)$. Any real-valued function of any random variable fits this structure, and therefore, is a random variable.

If $Y\big(X(\Omega)\big)$ is discrete, then Y is a discrete random variable. We can obtain its mass function as: $p_Y(y) = P_{Y(X(\Omega))}(\{y\}) = P_{X(\Omega)}(\{x \in X(\Omega): Y(x) = y\})$ for $y \in Y\big(X(\Omega)\big)$. Consider X to be uniformly distributed in $X(\Omega) = \{-2, -1, 0, 1, 2\}$. Then $p_X(x) = 1/5$ for all $x \in X(\Omega)$. Let $Y = X^2$. Then $Y\big(X(\Omega)\big) = \{0,1,4\}$, and the mass function of Y can be obtained as: $p_Y(0) = P_{X(\Omega)}(\{0\}) = p_X(0) = 1/5$, $p_Y(1) = P_{X(\Omega)}(\{-1,1\}) = p_X(-1) + p_X(1) = 2/5$, $p_Y(2) = P_{X(\Omega)}(\{-2,2\}) = p_X(-2) + p_X(2) = 2/5$. *Consider X to be uniformly distributed in $[-2,2]$ and $Y = \lfloor X \rfloor$, the floor function. Determine mass function of Y.*

Distribution function of Y, $F_Y(y) = P_{Y(X(\Omega))}\big((-\infty, y]\big) = P_{X(\Omega)}(\{x \in X(\Omega): Y(x) \le y\})$ for all $y \in \mathbb{R}$. If $F_Y(y)$ is continuous throughout in $\mathbb{R}$, then Y is a continuous random variable. Let $X \sim U(0,4)$ and $Y = \sqrt{X}$. Then $Y\big(X(\Omega)\big) = [0,2]$, and thus, $F_Y(y) = 0$ for $y < 0$ and 1 for $y > 2$. For $y \in [0,2]$, $F_Y(y) = P_{X(\Omega)}\big(\{x \in [0,4]: \sqrt{x} \le y\}\big) = P_{X(\Omega)}([0, y^2]) = y^2/4$. It's a continuous distribution function. Its density function is: $f_Y(y) = y/2$ for $y \in [0,2]$ and 0 otherwise. *Now, consider $Y = 1/X$. Obtain its distribution and density functions.*

For monotonic functions, we have a generic formula for obtaining distribution function. This is because monotonic functions are invertible, and the inverse functions are monotonic too. This simplifies the task of identifying $\{x \in X(\Omega): Y(x) \le y\}$, which is a key step in obtaining $F_Y(y)$. Consider $X(\Omega) = [0, \infty)$ and $Y(x) = x^2$. Y is a strictly increasing function and so is its inverse $Y^{-1}(y) = \sqrt{y}$. Here, $\{x \in [0,\infty): x^2 \le y\} = \{x \in [0, \infty): x \le \sqrt{y}\} = [0, \sqrt{y}]$. In general, for strictly increasing Y, Y^{-1} is strictly increasing too. Then $\{x \in X(\Omega): Y(x) \le y\} = \{x \in X(\Omega): x \le Y^{-1}(y)\}$, and thus, $F_Y(y) = P_{X(\Omega)}\big((-\infty, Y^{-1}(y)]\big) = F_X\big(Y^{-1}(y)\big)$.

Consider $X(\Omega) = (0, \infty)$ and $Y(x) = 1/x$. Y is a strictly decreasing function and so is its inverse $Y^{-1}(y) = 1/y$. Here, $\{x \in (0, \infty): 1/x \leq y\} = \{x \in (0, \infty): x \geq 1/y\} = [1/y, \infty)$. In general, for strictly decreasing Y, Y^{-1} is strictly decreasing too. Then $\{x \in X(\Omega): Y(x) \leq y\} = \{x \in X(\Omega): x \geq Y^{-1}(y)\}$. Note the reversal of inequality. At $x = Y^{-1}(y)$, $Y(x) = y$. Since Y is strictly decreasing in x, $Y(x) < y$ for all $x > Y^{-1}(y)$ and $Y(x) > y$ for all $x < Y^{-1}(y)$. Therefore, $Y(x) \leq y \equiv x \geq Y^{-1}(y)$, and thus, $F_Y(y) = P_{X(\Omega)}\big([Y^{-1}(y), \infty)\big) = 1 - P_{X(\Omega)}\Big(\big(-\infty, Y^{-1}(y)\big)\Big) = 1 - F_X\Big(\big(Y^{-1}(y)\big)^-\Big)$. Note the left-limit. If Y is continuous random variable, then the left-limit is same as $F_X\big(Y^{-1}(y)\big)$.

Among monotonic functions of a random variable, linear functions arise more often than others. Let $Y = aX + b$. If $a > 0$, then Y is strictly increasing, and if $a < 0$, Y is strictly decreasing. Sometimes, linear transformation retains nature of the probability distributions, e.g., normal distribution. Let $X \sim N(\mu, \sigma^2)$. Then $X(\Omega) = Y\big(X(\Omega)\big) = (-\infty, \infty)$ and

$$F_Y(y) = \begin{cases} F_X\big(Y^{-1}(y)\big) = F_X\left(\dfrac{y-b}{a}\right) = \displaystyle\int_{-\infty}^{(y-b)/a} f_X(x)dx \ \text{ if } a > 0 \\ 1 - F_X\Big(\big(Y^{-1}(y)\big)^-\Big) = 1 - F_X\left(\dfrac{y-b}{a}\right) = \displaystyle\int_{(y-b)/a}^{\infty} f_X(x)dx \ \text{ if } a < 0 \end{cases}$$

for all $y \in \mathbb{R}$. Using Leibniz integral rule, we obtain density function of Y as:

$$f_Y(y) = \begin{cases} \dfrac{1}{a} f_X\left(\dfrac{y-b}{a}\right) = \dfrac{1}{a\sqrt{2\pi\sigma^2}} e^{-\frac{((y-b)/a-\mu)^2}{2\sigma^2}} = \dfrac{1}{\sqrt{2\pi(a\sigma)^2}} e^{-\frac{(y-(a\mu+b))^2}{2(a\sigma)^2}} \ \text{ if } a > 0 \\ -\dfrac{1}{a} f_X\left(\dfrac{y-b}{a}\right) = \dfrac{1}{|a|\sqrt{2\pi\sigma^2}} e^{-\frac{((y-b)/a-\mu)^2}{2\sigma^2}} = \dfrac{1}{\sqrt{2\pi(|a|\sigma)^2}} e^{-\frac{(y-(a\mu+b))^2}{2(|a|\sigma)^2}} \ \text{ if } a < 0 \end{cases}$$

for all $y \in \mathbb{R}$. Observe that the density function of Y is same as that of a normal random variable with parameters $a\mu + b$ and $(|a|\sigma)^2$. If two random variables have the same density or distribution or mass functions, then they are identical. Therefore, $Y \sim N(a\mu + b, (|a|\sigma)^2)$. *Check if linear transformation retains uniform distribution.*

Let us consider the problem of obtaining distribution function value of $X \sim N(\mu, \sigma^2)$. We cannot do it analytically. Let us define $Z = (X - \mu)/\sigma$. Z is a linear transformation of X with $a = 1/\sigma > 0$ and $b = -\mu/\sigma$. Then $Z \sim N(0,1)$, the standard normal random variable. We denote distribution function of $N(0,1)$ by Φ and it is evaluated numerically. Then

$$F_X(x) = P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

for all $x \in \mathbb{R}$. If we want to calculate $F(88)$ or $F(121)$ for $N(100,15^2)$, then we have to find value of $\Phi((88-100)/15) = \Phi(-0.8)$ or $\Phi((121-100)/15) = \Phi(1.4)$. *Search normal distribution table and find these values*. Different types of tables are available. Most tables provide $\Phi(z)$ for $z \leq 0$. Since normal density function is symmetric about the mean, $\Phi(z) = 1 - \Phi(-z)$ for $z > 0$. Therefore, $\Phi(1.4) = 1 - \Phi(-1.4)$. Some tables provide $\Phi(z)$ for $z \geq$

0. Again, using symmetry, $\Phi(-z) = 1 - \Phi(z)$ for $z > 0$. Therefore $\Phi(-0.8) = 1 - \Phi(0.8)$. There are normal tables that provide values of $1 - \Phi(z)$ or $\Phi(z) - 0.5$ for $z \geq 0$. Familiarize yourself with different kinds of normal distribution tables. It is possible that your z value is not found in the table, then you need to use interpolation.

Expected value

Expected value or mean value of a random variable is the long-run average of the observed values of the random variable, when produced repeatedly. Let $x_1, x_2, \dots, x_n$ denote n such observed values of random variable X. Then the mean value of X, $E[X] = \lim_{n\to\infty} \sum_{i=1}^{n} x_i/n$, provided the limit exist. This quantity is equivalent to $\sum_{x\in X(\Omega)} xp(x)$ for the case of discrete random variables and $\int_{-\infty}^{\infty} xf(x)dx$ for the case of continuous random variables.

Let $X(\Omega) = \{v_1, v_2, \dots, v_m\}$. Each of the observed values of X, i.e., $x_1, x_2, \dots, x_n$, is one of the outcomes of $X(\Omega)$. Let n_j denote the number of times the j-th outcome, i.e., v_j, is observed. Then $\sum_{i=1}^{n} x_i = \sum_{j=1}^{m} v_j n_j$, and using the frequentists' measure, $E[X] = \lim_{n\to\infty} \sum_{i=1}^{n} x_i/n = \lim_{n\to\infty} \sum_{j=1}^{m} v_j n_j/n = \sum_{j=1}^{m} v_j \lim_{n\to\infty} n_j/n = \sum_{j=1}^{m} v_j P(X = v_j) \Rightarrow E[X] = \sum_{x\in X(\Omega)} xp(x)$ as claimed. If $X(\Omega)$ is infinite, then the exchange of limit and sum is permitted if and only if $\lim_{n\to\infty} \sum_{i=1}^{n} x_i/n$ exists (i.e., the limit converges), equivalently $\sum_{x\in X(\Omega)} xp(x)$ exists (i.e., the sum is finite). In the continuous case, $E[X] = \int_{-\infty}^{\infty} xf(x)dx$, provided it exists (i.e., the integral is finite). Proof for this case is rather difficult.

The above definition and the equivalence seem to suggest that there are random variables that do not have expected value. Consider $p(x) = 1/x(1+x)$ for $x = 1,2,3, \dots$. It is a valid mass function, because $p(x) > 0\ \forall x \in X(\Omega)$ and $\sum_{x=1}^{\infty} 1/x(1+x) = \sum_{x=1}^{\infty}(1/x - 1/(1+x)) = (1/1 - 1/2) + (1/2 - 1/3) + (1/3 - 1/4) + \cdots = 1$. Its expected value is: $\sum_{x=1}^{\infty} xp(x) = \sum_{x=1}^{\infty} 1/(1+x) = 1/2 + 1/3 + 1/4 + \cdots$, which does not exist (i.e., it diverges to ∞). Thus, the mean value does not exist in this case. *Consider $f(x) = 1/x^2$ for $x \geq 1$ and 0 otherwise. Check if it is a valid density and if its expectation exists.*

Now, we obtain expected value of the common random variables. First, we consider uniform random variable in $[a, b]$. Here, $f(x) = 1/(b-a)$ for $x \in [a, b]$, and 0 otherwise.

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_{a}^{b} \frac{x}{b-a} dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

Let us consider binomial random variable with parameters n (number of trials) and p (success probability). Here, $p(x) = C(n, x)p^x(1-p)^{n-x}$ for $x = 0,1, \dots, n$.

$$E[X] = \sum_{x\in X(\Omega)} xp(x) = \sum_{x=0}^{n} xC(n, x)p^x(1-p)^{n-x} = \sum_{x=1}^{n} \frac{xn!}{x!\,(n-x)!} p^x(1-p)^{n-x}$$

$$= np\sum_{x=1}^{n}\frac{(n-1)!}{(x-1)!\,(n-x)!}p^{x-1}(1-p)^{n-x} = np\sum_{y=0}^{m}C(m,y)p^{y}(1-p)^{m-y} = np$$

by replacing $x-1=y$ and $n-1=m$. Note that the last summands are mass function of a binomial random variable with parameters m and p; hence, the sum is 1.

Now we consider geometric random variable with parameter p (success probability). Let $q = 1-p$. Then, $p(x) = (1-p)^{x-1}p = q^{x-1}p$ for $x = 1,2,3,\ldots$.

$$E[X] = \sum_{x\in X(\Omega)} xp(x) = \sum_{x=1}^{\infty} xq^{x-1}p = p\sum_{x=1}^{\infty}\frac{dq^x}{dq} = p\frac{d}{dq}\left(\sum_{x=1}^{\infty}q^x\right) = p\frac{d}{dq}\left(\frac{q}{1-q}\right)$$
$$= p\frac{(1-q)1-q(-1)}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

Let us consider Poisson random variable with parameters λ (intensity) and t (duration). Here, $p(x) = e^{-\lambda t}\,(\lambda t)^x/x!$ for $x = 0,1,2,\ldots$.

$$E[X] = \sum_{x\in X(\Omega)} xp(x) = \sum_{x=0}^{\infty} xe^{-\lambda t}\frac{(\lambda t)^x}{x!} = \lambda t\sum_{x=1}^{\infty}e^{-\lambda t}\frac{(\lambda t)^{x-1}}{(x-1)!} = \lambda t\sum_{y=0}^{\infty}e^{-\lambda t}\frac{(\lambda t)^y}{y!} = \lambda t$$

by replacing $x-1=y$. Note that the last summands are mass function of the given Poisson random variable; hence, the sum is 1. If the duration $t=1$, then $E[X]=\lambda$, and for any other duration, expectation is t times the rate λ.

Now we consider exponential random variable with parameter λ (rate). Here, $f(x) = \lambda e^{-\lambda x}$ for $x > 0$ and 0 otherwise.

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_0^{\infty} x\lambda e^{-\lambda x}dx = \left[x\frac{e^{-\lambda x}}{-1}\right]_0^{\infty} - \int_0^{\infty} 1\frac{e^{-\lambda x}}{-1}dx \ \text{(integration by parts)}$$
$$= \left[\frac{x}{e^{\lambda x}}\right]_{\infty}^{0} + \int_0^{\infty} e^{-\lambda x}dx = \left(\frac{0}{e^0} - \lim_{x\to\infty}\frac{x}{e^{\lambda x}}\right) + \left[\frac{e^{-\lambda x}}{-\lambda}\right]_0^{\infty}$$
$$= \left(0 - \lim_{x\to\infty}\frac{1}{\lambda e^{\lambda x}}\right) + \left[\frac{e^{-\lambda x}}{-\lambda}\right]_0^{\infty} \ \text{(L'Hopital's rule)}$$
$$= 0 + \frac{1}{\lambda}\left[\frac{1}{e^{\lambda x}}\right]_{\infty}^{0} = \frac{1}{\lambda}\left(\frac{1}{e^0} - \lim_{x\to\infty}\frac{1}{e^{\lambda x}}\right) = \frac{1}{\lambda}$$

Finally, let us consider normal random variable with parameters μ and σ^2. We do not know meanings of μ and σ^2. Here, $f(x) = \left(1/\sqrt{2\pi\sigma^2}\right)\exp(-(x-\mu)^2/2\sigma^2)$ for all $x \in \mathbb{R}$.

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty}\frac{x}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx = \int_{-\infty}^{\infty}\frac{\mu+z\sigma}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz \ \text{replacing}\ \frac{x-\mu}{\sigma} = z$$

$$= \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-\frac{z^2}{2}} dz = \mu \times 1 + \frac{\sigma}{\sqrt{2\pi}} \times 0 = \mu$$

Note that the first integrand is density function of standard normal random variable (where $\mu = 0$ and $\sigma^2 = 1$); hence, the integration is 1. The second integrand is odd function; hence, the integration is 0. So, the parameter μ of normal distribution is its mean value.

Let us consider a real-valued function Y of random variable X. We know $Y(X)$ is a random variable, and thus, we can compute $E[Y]$ by employing: $\sum_{y \in Y(X(\Omega))} y p_Y(y)$ if Y is discrete or $\int_{-\infty}^{\infty} y f_Y(y) dy$ if Y is continuous. For this, we need to obtain mass/density function of Y first. However, there is a simpler way, referred to as the law of the unconscious statistician, that does not require the knowledge of mass/density of Y. Since X is the source of uncertainty in Y, the same can be used to obtain expectation of Y. Then $E[Y] = \sum_{x \in X(\Omega)} Y(x) p_X(x)$ if X is discrete or $E[Y] = \int_{-\infty}^{\infty} Y(x) f_X(x) dx$ if X is continuous.

Consider X to be discrete. Then Y is discrete as well, and $p_Y(y) = P_{X(\Omega)}(\{x \in X(\Omega): Y(x) = y\})$, which can be expressed as follows using the third axiom:

$$p_Y(y) = P_{X(\Omega)}\left(\bigcup_{\substack{x \in X(\Omega) \\ Y(x)=y}} \{x\}\right) = \sum_{\substack{x \in X(\Omega) \\ Y(x)=y}} P_{X(\Omega)}(\{x\}) = \sum_{x \in X(\Omega)} p_X(x) \cdot \mathbb{1}(Y(x) = y)$$

where $\mathbb{1}(Y(x) = y)$ is the indicator function for the condition $Y(x) = y$. It that takes value 1 if the condition holds, else it is 0. Then $E[Y] = \sum_{y \in Y(X(\Omega))} y p_Y(y)$ can be written as:

$$E[Y] = \sum_{y \in Y(X(\Omega))} y \sum_{x \in X(\Omega)} p_X(x) \cdot \mathbb{1}(Y(x) = y) = \sum_{x \in X(\Omega)} p_X(x) \sum_{y \in Y(X(\Omega))} y \cdot \mathbb{1}(Y(x) = y)$$

For a given $x \in X(\Omega)$, $Y(x)$ is unique and it is in $Y\big(\mathrm{X}(\Omega)\big)$, and hence, $\mathbb{1}(Y(x) = y) = 1$ for exactly one $y \in Y\big(X(\Omega)\big)$. So, the inner sum reduces to $Y(x)$ for every $x \in X(\Omega)$, and then $E[Y] = \sum_{x \in X(\Omega)} Y(x) p_X(x)$, as claimed. The above argument can be better understood with an example. Proof for the other cases is more involved.

Let us apply the law of the unconscious statistician to the linear transformation of a random variable. Let $Y = aX + b$. If X is discrete, then $E[Y] = \sum_{x \in X(\Omega)} Y(x) p_X(x)$ simplifies to

$$E[Y] = \sum_{x \in X(\Omega)} (ax + b) p_X(x) = a \sum_{x \in X(\Omega)} x p_X(x) + b \sum_{x \in X(\Omega)} p_X(x) = aE[X] + b.$$

Show the above for continuous X. In a similar manner, one can show that $E[\sum_{i=1}^{n} Y_i(X)] = \sum_{i=1}^{n} E[Y_i(X)]$ for $n \geq 1$ where $Y_1, Y_2, Y_3, \dots$ are real-valued functions of X.

With the help of the law of the unconscious statistician, we can obtain higher moments of a random variable. k-th moment, for $k = 1,2,3, \dots$, of random variable X is defined as $E[X^k]$. The expectation is the first moment. We can consider X^k as a real-valued function of X, and then $E[X^k] = \sum_{x \in X(\Omega)} x^k p_X(x)$ if X is discrete or $\int_{-\infty}^{\infty} x^k f_X(x) dx$ if X is continuous. Let us consider another real-valued function e^{sX} where $s \in \mathbb{R}$ is so chosen that $E[e^{sX}]$ exists. We denote the expectation by $m_X(s)$ and refer to it as the moment generating function of X. The reason for such naming is that $m_X(s)$ contain information about the moments of X.

$$m_X(s) = E[e^{sX}] = E\left[1 + sX + \frac{(sX)^2}{2!} + \frac{(sX)^3}{3!} + \cdots\right] \text{ by infinite series expansion}$$
$$= 1 + sE[X] + \frac{s^2}{2!}E[X^2] + \frac{s^3}{3!}E[X^3] + \cdots$$
$$\Rightarrow m_X^{(k)}(s) = \frac{d^k m_X(s)}{ds^k} = E[X^k] + \frac{s}{1!}E[X^{k+1}] + \frac{s^2}{2!}E[X^{k+2}] + \cdots \text{ for } k = 1,2,3, \ldots$$

Assume $m_X(s)$ is defined for some interval around $s = 0$. Then its k-th derivative at $s = 0$, $m_X^{(k)}(0) = E[X^k]$ is the k-th moment of X, for $k = 1,2,3, \ldots$. The actual purpose of moment generating function is not the moments. It is a way of capturing randomness, like distribution or mass or density functions do. If two random variables have the same moment generating function in some interval around $s = 0$, then their distribution and mass/density functions are the same. This property will be useful later.

Variance and quantiles

Variance of a random variable is the long-run average deviation of the observed values, when produced repeatedly, from its expected value. The deviation is measured in squared distance. Let $x_1, x_2, \ldots, x_n$ denote n such observed values of random variable X. Then the variance of X, $Var(X) = \lim_{n\to\infty} \sum_{i=1}^{n} (x_i - E[X])^2/n$, provided the limit exist. Let $Y = (X - E[X])^2$, a real-valued function of X. Then $Var(X) = E[Y]$, i.e.,

$$Var(X) = E[(X - E[X])^2] = E[X^2 - 2XE[X] + (E[X])^2]; \text{ note that } E[X] \text{ is a constant}$$
$$= E[X^2] - 2E[X]E[X] + E^2[X] = E[X^2] - E^2[X]$$

Note that first two moments of a random variable must exist for its variance to exist. By the definition, variance is negative, and its positive square root is known as standard deviation. Now, we obtain variance of the common random variables. We start with uniform random variable in $[a, b]$. We already obtained $E[X] = (a + b)/2$. Then

$$Var(X) = E[X^2] - E^2[X] = \int_a^b \frac{x^2}{b-a} dx - E^2[X] = \frac{b^3 - a^3}{3(b-a)} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

Let us consider binomial random variable with parameters n (number of trials) and p (success probability). We already obtained $E[X] = np$. Now, let us obtain $E[X(X-1)]$.

$$E[X(X-1)] = \sum_{x \in X(\Omega)} x(x-1)p(x) = \sum_{x=0}^{n} x(x-1)C(n,x)p^x(1-p)^{n-x}$$

$$= \sum_{x=2}^{n} \frac{x(x-1)n!}{x!\,(n-x)!} p^x (1-p)^{n-x} = n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!\,(n-x)!} p^{x-2}(1-p)^{n-x}$$

$$= n(n-1)p^2 \sum_{y=0}^{m} C(m,y)p^y(1-p)^{m-y} = n(n-1)p^2 \text{ by } x-2=y \text{ and } n-2=m$$

Note that the last summands are mass function of binomial random variable with parameters m and p; thus, the sum is 1. Since $E[X(X-1)] = E[X^2] - E[X]$, then $Var(X) = E[X^2] - E^2[X] = E[X(X-1)] + E[X] - E^2[X] = n(n-1)p^2 + np - (np)^2 = np(1-p)$.

Now we consider geometric random variable with parameter p (success probability). Let $q = 1-p$. We already obtained $E[X] = 1/p$. Let us obtain $E[X(X-1)]$ now.

$$E[X(X-1)] = \sum_{x \in X(\Omega)} x(x-1)p(x) = \sum_{x=1}^{\infty} x(x-1)q^{x-1}p = p\sum_{x=2}^{\infty}(x-1)\frac{dq^x}{dq}$$

$$= p\frac{d}{dq}\left(\sum_{x=2}^{\infty}(x-1)q^x\right) = p\frac{d}{dq}\left(q^2\sum_{x=2}^{\infty}\frac{dq^{x-1}}{dq}\right) = p\frac{d}{dq}\left(q^2\frac{d}{dq}\left(\sum_{x=2}^{\infty}q^{x-1}\right)\right)$$

$$= p\frac{d}{dq}\left(q^2\frac{d}{dq}\left(\frac{q}{1-q}\right)\right) = p\frac{d}{dq}\left(\frac{q^2}{(1-q)^2}\right) = \frac{2pq}{(1-q)^3} = \frac{2(1-p)}{p^2}$$

Since $E[X(X-1)] = E[X^2] - E[X]$, $Var(X) = E[X^2] - E^2[X] = E[X(X-1)] + E[X] - E^2[X] = 2(1-p)/p^2 + 1/p - 1/p^2 = (1-p)/p^2$.

Let us consider Poisson random variable with parameters λ (intensity) and t (duration). We already obtained $E[X] = \lambda t$. Again, we obtain $E[X(X-1)]$.

$$E[X(X-1)] = \sum_{x \in X(\Omega)} x(x-1)p(x) = \sum_{x=0}^{\infty} x(x-1)e^{-\lambda t}\frac{(\lambda t)^x}{x!} = (\lambda t)^2\sum_{x=2}^{\infty} e^{-\lambda t}\frac{(\lambda t)^{x-2}}{(x-2)!}$$

$$= (\lambda t)^2\sum_{y=2}^{\infty} e^{-\lambda t}\frac{(\lambda t)^y}{y!} = (\lambda t)^2 \text{ by } x-2=y$$

The last summands are mass function of the given Poisson random variable; so, the sum is 1. Since $E[X(X-1)] = E[X^2] - E[X]$, $Var(X) = E[X^2] - E^2[X] = E[X(X-1)] + E[X] - E^2[X] = (\lambda t)^2 + \lambda t - (\lambda t)^2 = \lambda t$, which is same as the expectation.

Now we consider exponential random variable with parameter λ (rate). We already obtained $E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_0^{\infty} x\lambda e^{-\lambda x}dx = 1/\lambda$. Then

$$Var(X) = E[X^2] - E^2[X] = \int_0^{\infty} x^2\lambda e^{-\lambda x}dx - \frac{1}{\lambda^2} = \left[x^2\frac{e^{-\lambda x}}{-1}\right]_0^{\infty} - \int_0^{\infty} 2x\frac{e^{-\lambda x}}{-1}dx - \frac{1}{\lambda^2}$$

$$= \left(-\lim_{x\to\infty}\frac{x^2}{e^{\lambda x}} + \frac{0}{e^0}\right) + \frac{2}{\lambda}\int_0^{\infty} x\lambda e^{-\lambda x}dx - \frac{1}{\lambda^2} = \left(-\lim_{x\to\infty}\frac{2}{\lambda^2 e^{\lambda x}}\right) + \frac{2}{\lambda}\cdot\frac{1}{\lambda} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Finally, let us consider normal random variable with parameters μ (mean) and σ^2. We do not know the meanings of σ^2 yet. We already obtained $E[X] = \mu$. Now,

$$Var(X) = E[(X - E[X])^2] = \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{\sigma^2 z^2}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \text{ by } \frac{x-\mu}{\sigma} = z$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}} dz = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z g(z) dz \text{ where } g(z) = z e^{-\frac{z^2}{2}}$$

Now $\int g(z)dz = \int z e^{-\frac{z^2}{2}} dz = \int e^{-y} dy = -e^{-y} = -e^{-\frac{z^2}{2}}$ by $\frac{z^2}{2} = y$

Then $\int_{-\infty}^{\infty} z g(z) dz = \left[-z e^{-\frac{z^2}{2}}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz$ (integration by parts)

$$= \left(-\lim_{z\to\infty} \frac{z}{e^{z^2/2}} + \lim_{z\to-\infty} \frac{z}{e^{z^2/2}}\right) + \sqrt{2\pi} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \text{ (integrand is density of } N(0,1))$$

$$= \left(\lim_{z\to-\infty} \frac{1}{z e^{z^2/2}} - \lim_{z\to\infty} \frac{1}{z e^{z^2/2}}\right) + \sqrt{2\pi} = \sqrt{2\pi} \Rightarrow Var(X) = \frac{\sigma^2}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = \sigma^2$$

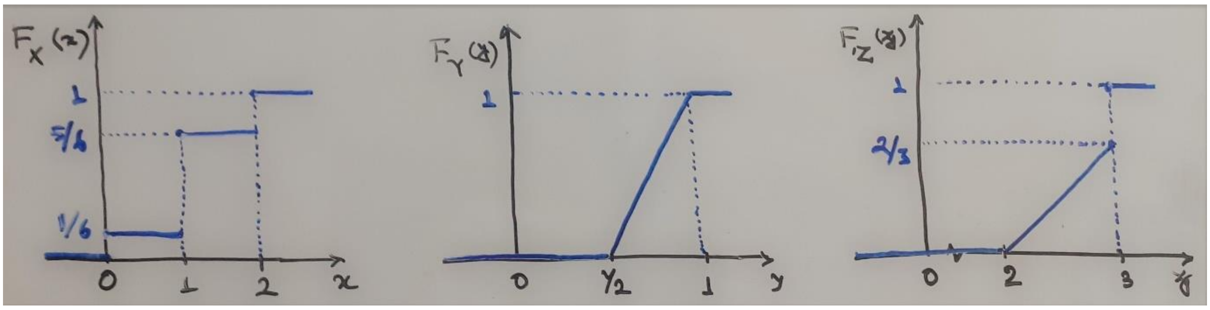
So, the parameter σ^2 of normal distribution is its variance.

Let us obtain variance of the linear transformation of a random variable. Let $Y = aX + b$. We already obtained $E[Y] = aE[X] + b$ by the law of unconscious statistician. Then

$$\begin{aligned} Var(Y) &= E[Y^2] - E^2[Y] = E[(aX+b)^2] - E^2[Y] \\ &= E[a^2X^2 + 2abX + b^2] - (aE[X] + b)^2 \\ &= (a^2E[X^2] + 2abE[X] + b^2) - (a^2E^2[X] + 2abE[X] + b^2) \\ &= a^2(E[X^2] - E^2[X]) = a^2 Var(X) \end{aligned}$$

Observe the effect of a on $Var(Y)$ is much more pronounced than its effect on $E[Y]$. On the other hand, b has no effect on $Var(Y)$.

Before we conclude this module let us discuss quantiles, which is used for simulating any random variable. We have learned to obtain distribution function $F(x)$ at a given $x \in \mathbb{R}$ for any random variable. Quantile is inverse of the distribution function. The q-th quantile for a given $q \in (0,1)$ is: $x_q \in \mathbb{R}$ such that $F(x_q) = q$. Consider the distribution: $F_Y(y) = 0$ for $y < 0.5$, $2y - 1$ for $y \in [0.5,1]$, and 1 for $y > 1$. It's the 2nd diagram below. For $q \in (0,1)$, $F_Y(y_q) = q$ gives $2y_q - 1 = q \Rightarrow y_q = (1+q)/2$ as the q-th quantile.



Now, consider the distribution: $F_Z(z) = 0$ for $z < 2$, $(z-2)/1.5$ for $z \in [2,3)$, and 1 for $z \geq 3$. It's the 3rd diagram above. Consider $q = 3/4$. We see that there is no $z_q \in \mathbb{R}$ such that $F_Z(z_q) = 3/4$. This happens because $F_Z(z)$ jumps from $2/3$ to 1 at $z = 3$. It makes sense to select $z_{3/4} = 3$, because $F_Z(3^-) = 2/3 < 3/4 < 1 = F_Z(3)$. In general, if we select $x_q = \min\{x \in \mathbb{R}: F(x) \geq q\}$, instead of solving $F(x_q) = q$, the problem of jumps in distribution function is resolved. This also provides unique quantile value when $F(x_q) = q$ has multiple solutions. Consider the distribution: $F_X(x) = 0$ for $x < 0$, $1/6$ for $x \in [0,1)$, $5/6$ for $x \in [1,2)$, and 1 for $x \geq 2$. It's the 1st diagram above. Note that all $x_q \in [1,2)$ solves $F_X(x_q) = 5/6$, but $x_q = \min\{x \in \mathbb{R}: F(x) \geq 5/6\} = 1$.

At times, we are interested in complex functions of a random variable. Let $Y(X)$ denote such a complex function and we are interested in obtaining $E[Y]$, or some such quantity, but we cannot do it analytically. Then a way out is to conduct Ω repeatedly and observe the values of X, denoted by $x_1, x_2, \ldots, x_n$. Let $y_i = Y(x_i)$ for $i = 1,2,\ldots,n$ denote the corresponding values of Y. Then $E[Y] \approx \sum_{i=1}^{n} y_i/n$, and this approximation gets better as n increases. One problem with this approach is that performing Ω repeatedly can be costly or even impossible. Then we simulate X, instead of producing it from Ω.

Modern computers have random number generator that produces numbers in between 0 and 1 such that all outcomes are (more-or-less) equally likely. This takes care of $X \sim U(0,1)$. Even if $X \sim U(a,b)$, we can use $Q \sim U(0,1)$ to obtain $X = a + Q(b-a)$ that follows $U(a,b)$. In order to simulate any kind of random variable, we need to treat $Q \sim U(0,1)$ generated by the computer as the distribution function value of the random variable. All distribution functions take values in $[0,1]$. Conversion from $Q \in (0,1)$ to an observed value of the random variable happens through quantile x_Q, which is taken as the observed value. From the definition of quantile, $Q \leq F(x) \equiv x_Q \leq x$. Then $P(x_Q \leq x) = P(Q \leq F(x)) = F(x)$ as $Q \sim U(0,1)$. So, this approach produces observations according to the given distribution.

Practice problems

Book: A Modern Introduction to Probability and Statistics by Dekking et al.

Discrete random variables
Chapter-4, Exercise No. 4, 6, 12, 13, 14

Continuous random variables
Chapter-5, Exercise No. 3, 5, 7, 9, 13

Functions of random variables
Chapter-8, Exercise No. 3, 4, 5, 6, 8, 10

Mean, variance, and quantiles
Chapter-7, Exercise No. 7, 8, 14
Chapter-5, Exercise No. 11, 14