MA2202: Probability I

Random variables

Spring 2021

Satvik Saha
19MS154

Indian Institute of Science Education and Research, Kolkata,
Mohanpur, West Bengal, 741246, India.

Definition 3.1 (Random variable)**.** Given a probability space $(\Omega, \mathcal{E}, P)$, a function $X : \Omega \to \mathbb{R}$ is called a random variable if $X^{-1}(r, \infty) \in \mathcal{E}$ for all $r \in \mathbb{R}$.

Remark. For some $S \subseteq \mathbb{R}$, we denote

$$P(X \in S) = P(\{\omega \in \Omega \colon X(\omega) \in S\}).$$

Definition 3.2 (Discrete random variable)**.** A random variable which can assume only a countably infinite number of values is called a discrete random variable.

Example. Let $X : \Omega \to \mathbb{R}$ denote the number of heads obtained when a fair coin is tossed thrice. Note that $\Omega = \{0, 1, 2, 3\}$. Thus, $P(X = 0) = P(X = 4) = 1/8$ and $P(X = 1) = P(X = 2) = 3/8$.

Definition 3.3 (Probability distribution)**.** The probability distribution of a random variable X is the set of pairs $(X(A), P(A))$ for all $A \in \mathcal{E}$.

Definition 3.4 (Probability mass function)**.** Let X be a discrete random variable. The probability mass function of X is the function $p_X : \mathbb{R} \to [0, 1]$,

$$p_X(\alpha) = P(X = \alpha).$$

Remark. Since X is a discrete random variable, the set $S = \{x \in \mathbb{R} \colon p_X(\alpha) > 0\}$ is countable, and

$$\sum_{x \in S} p_X(x) = 1.$$

Definition 3.5 (Expectation)**.** The expectation of $g(X)$, for $g\colon \mathbb{R} \to \mathbb{R}$ and a discrete random variable X is defined as

$$E\left[g(X)\right] = \sum_{x\in S} g(x)\, p_X(x),$$

if the series converges absolutely.

Example. The n^{th} moment of a discrete random variable $E\left[X^n\right]$ is defined as

$$E\left[X^n\right] = \sum_{x\in S} x^n\, p_X(x),$$

if the series converges absolutely.

The first moment $\mu = E\left[X\right]$ is called the mean. The second moment $\sigma^2 = E\left[(X-\mu)^2\right]$ is called the variance. Note that

$$\sigma^2 = \sum (x-\mu)^2\, p(x) = \sum x^2\, p(x) - 2\mu x\, p(x) + \mu^2 p(x).$$

Simplifying,

$$\sigma^2 = E\left[X^2\right] - E\left[X\right]^2.$$

Definition 3.6 (Cumulative distribution function)**.** The cumulative distribution function of a random variable X is defined as the function $F_X\colon \mathbb{R} \to [0,1]$,

$$F_X(\alpha) = P(X \le \alpha).$$

Definition 3.7 (Continuous random variable)**.** A continuous random variable X is such that its cumulative distribution function F_X is continuous.

Definition 3.8 (Probability density function)**.** Let X be a continuous random variable with a cumulative distribution function F_X. If we write

$$F_X(\alpha) = \int_{-\infty}^{\alpha} f_X(x)\, dx$$

for all $\alpha \in \mathbb{R}$, then f_X is a probability density function. If f_X is continuous, then the Fundamental Theorem of Calculus guarantees that $f_X(x) = F_X'(x)$.

Remark. Note that we can write

$$P(\alpha \le X \le \beta) = \int_{\alpha}^{\beta} f_X(x)\, dx.$$

We also demand

$$\int_{-\infty}^{+\infty} f_X(x)\, dx = 1.$$

Example. The uniform distribution on an interval $[a, b] \subset \mathbb{R}$ is defined using the probability density function

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

Definition 3.9 (Expectation)**.** The expectation of $g(X)$, for $g\colon \mathbb{R} \to \mathbb{R}$ and a continuous random variable X is defined as

$$E\left[g(X)\right] = \int_{-\infty}^{+\infty} g(X)\, f_X(x)\, dx$$

if the integral converges absolutely.

Definition 3.10 (Mixed random variable)**.** A random variable whose cumulative distribution function F_X is discontinuous at countably many points, with F_X being continuous and strictly increasing in at least one interval is called a mixed random variable.

Definition 3.11 (Conditional probability distributions)**.** Let X and Y be two random variables, and let $A, B \subseteq \mathbb{R}$. If $P(Y \in B) > 0$, then

$$P(X \in A \,|\, Y \in B)\, P(Y \in B) = P(X \in A, Y \in B).$$

Definition 3.12 (Independent random variables)**.** We say that X and Y are independent if

$$P(X \in A, Y \in B) = P(X \in A)\, P(Y \in B)$$

for all $A, B \subseteq \mathbb{R}$.

Example. Consider an experiment where a fair die is rolled until a 6 is obtained, and let X be the random variable denoting the number of throws. Also, let Y be a random variable which is 1 if the first even outcome is a 6, and 0 otherwise. Observe that $P(X = 1) = 1/6$ and $P(Y = 1) = 1/3$. Also, $P(X = 1, Y = 1) = 1/6 \neq P(X = 1)\, P(Y = 1)$, hence X and Y are not independent random variables.

It can be shown that

$$P(X = n \,|\, Y = 1) = \frac{1}{2^n}.$$

Bernoulli distribution

Consider an experiment with one trial, which has two possible outcomes; the probability of a success is given by p and the probability of a failure is $q = 1 - p$. The discrete random variable such that $X = 1$ on success and $X = 0$ on failure is said to follow the Bernoulli distribution. Note that the expectation value is simply

$$E\left[X\right] = p.$$

Binomial distribution

Consider an experiment with n Bernoulli trials. We could let X_i be a random variable denoting the outcome of the i^{th} trial, so we demand that $\{X_i\}$ are independent and identically distributed. The distribution of the sum $X = X_1 + \cdots + X_n$ follows the Binomial distribution $B(n, p)$, where

$$P(X = k) = \binom{n}{k} p^k q^{n-k}.$$

The expectation value is given by

$$E[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = np.$$

This is more easily seen from the linearity of expectation,

$$E[X] = \sum_{i=1}^{n} E[X_i] = np.$$

Geometric distribution

Consider an experiment where Bernoulli trials are repeated until success. The random variable denoting the number of trials required is said to follow the geometric distribution, where the probability that n trials were required is given by

$$P(X = n) = pq^{n-1}.$$

Pascal distribution

Consider an experiment where Bernoulli trials are repeated until k successes. The random variable denoting the number of trials required follows the Pascal, or negative binomial distribution. The probability that $n \geq k$ trials were required is given by

$$P(X = n) = \binom{n-1}{k-1} p^k q^{n-k}.$$

Note that we choose $k - 1$ successes from $n - 1$ trials, since the last trial is by definition a success.

Remark. This can be derived by noting that if $X_i \sim \text{Geometric}(p)$, then the random variable $X \sim \text{Pascal}(k, p)$ is simply

$$X = X_1 + X_2 + \cdots + X_k,$$

so for success in n trials, we find $P(X = n)$. This means that we need to sum the probabilities for all possible $X_i = n_i$, where $n = \sum n_i$. There are $\binom{n-1}{k-1}$ such solutions, and each of these solutions must occur with probability $p^k q^{n-k}$, denoting k successes and the remaining $n - k$ failures.

Hypergeometric distribution

Consider an experiment where we choose n balls randomly from a population of N distinct balls, of which m are red. The random variable denoting the number of red balls follows the hypergeometric distribution. The probability of obtaining $k \leq m$ red balls is given by

$$P(X = k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}.$$

Poisson distribution

The probability mass function of the Poisson distribution is given by

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

where $k \in \mathbb{N}$. This is generally used to model the number of times an event occurs in a given interval. This is a limiting form of the binomial distribution. Note that the expectation value is given by

$$E[X] = \sum_{k=0}^{\infty} k \cdot \frac{\lambda^k}{k!} e^{-\lambda} = \lambda.$$

In addition, the variance is also given by $\sigma^2 = \lambda$.

Exponential distribution

The probability density function of the exponential distribution is given by

$$f_X(x) = \lambda e^{-\lambda x},$$

where $x \geq 0$ and 0 elsewhere. This is generally used to model the waiting time between successive events. Note that the expectation value is given by

$$E[X] = \int_0^{\infty} \lambda x e^{-\lambda x}\, dx = \frac{1}{\lambda},$$

In addition, the variance is given by $\sigma^2 = 1/\lambda^2$.

Definition 3.13 (Memorylessness)**.** The probability distribution of a non-negative discrete random variable X is called memoryless if for all $m, n \in \mathbb{Z}_{\geq 0}$,

$$P(X > m + n \,|\, X > n) = P(X > m).$$

Remark. This can be written as

$$P(X > m + n) = P(X > m)\, P(X > n).$$

Again, a continuous random variable Y is called memoryless if for all $s, t \geq 0$,

$$P(X > s + t \,|\, X > t) = P(X > s).$$

Example. Suppose that $X \sim \text{Geometric}(p)$. Then,

$$P(X > n) = \sum_{k=n+1}^{\infty} pq^{k-1} = q^n,$$

where $q = 1 - p$. Also,

$$P(X > m + n, X > n) = P(X > m + n) = q^{m+n},$$

which gives

$$P(X > m + n, X > n) = q^m \cdot q^n = P(X > m)\, P(X > n).$$

Example. Suppose that $Y \sim \text{Exponential}(\lambda)$. Then,

$$P(Y > s) = \int_s^\infty \lambda e^{-\lambda t}\, dt = e^{-\lambda s}.$$

Thus,

$$P(Y > s + t) = P(Y > s)\, P(Y > t).$$

Normal distribution

The probability density function of the normal distribution is given by

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}.$$

Exercise 3.1. Consider a continuous random variable with probability density function

$$f(x) = ce^{-x^2/2}.$$

Determine the constant c.

Solution. Define

$$I = \int_{-\infty}^{+\infty} e^{-x^2/2}\, dx.$$

By using two dummy variables x and y, we can write

$$I^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(x^2+y^2)/2}\, dx\, dy.$$

We can shift to polar coordinates by using the transformations

$$x = r\cos\theta, \qquad y = r\sin\theta, \qquad dx\, dy = r\, d\theta\, dr.$$

Thus,

$$I^2 = \int_0^\infty \int_0^{2\pi} re^{-r^2/2}\, d\theta\, dr = \int_0^{2\pi} d\theta \int_0^\infty re^{-r^2/2}\, dr.$$

Now, the second integral can be computed as

$$\int_0^\infty re^{-r^2/2}\, dr = \int_0^\infty e^{-r^2/2}\, d(r^2/2) = -e^{-r^2/2}\Big|_0^\infty = 1.$$

Thus, we have

$$I = \int_{-\infty}^{+\infty} e^{-x^2/2}\, dx = \sqrt{2\pi}.$$

This is called a Gaussian integral. We must thus have $c = 1/\sqrt{2\pi}$.

Exercise 3.2. Find the moments of the standard normal distribution.

Solution. Note that $E\left[X^{2n-1}\right] = 0$, because the integral

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^{2n-1} e^{-x^2/2}\, dx = 0.$$

The function we are integrating over is an odd function. Now, we consider the second moment. Compute

$$\int_{-\infty}^{+\infty} x^2 e^{-x^2/2}\, dx = 2\int_0^{\infty} x^2 e^{-x^2/2}\, dx = -2xe^{-x^2/2}\Big|_0^{\infty} + 2\int_0^{\infty} e^{-x^2/2}\, dx = \sqrt{2\pi}.$$

Thus, $E\left[X^2\right] = 1$, which means that the variance is also 1. For higher moments, note that

$$\int x^{2n} e^{-x^2/2} = -x^{2n-1}e^{-x^2/2} + (2n-1)\int x^{2n-2} e^{-x^2/2}\, dx,$$

which gives the recurrence

$$E\left[X^{2n}\right] = (2n-1)\, E\left[X^{2n-2}\right] = (2n-1)!!.$$

Suppose that $Z \sim \text{Normal}(\mu, \sigma^2)$, note that $(Z-\mu)/\sigma$ is a standard normal variable. The cumulative distribution function of a standard normal variable is

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{x} e^{-t^2/2}\, dt.$$

Thus,

$$P(Z \le z) = P\left(\frac{Z-\mu}{\sigma} \le \frac{z-\mu}{\sigma}\right) = \Phi\left(\frac{z-\mu}{\sigma}\right).$$

Thus, we see that

$$E\left[(Z-\mu)^n\right] = \begin{cases} 0, & \text{if } n \text{ is odd} \\ \sigma^n (n-1)!!, & \text{if } n \text{ is even} \end{cases}.$$

Theorem 3.1 (Markov's inequality)**.** *Let X be a non-negative random variable. Then for all $a > 0$,*

$$P(X \ge a) \le \frac{E[X]}{a}.$$

Proof. Let $u_a \colon \mathbb{R}_{\ge 0} \to \{0, 1\}$ be a step function where $u_a(x) = 1$ when $x \ge a$ and 0 otherwise. Note that

$$u_a(x) \le \frac{x}{a}.$$

Then,

$$P(X \ge a) = E\left[u_a(X)\right] \le E\left[\frac{X}{a}\right] = \frac{E[X]}{a}.$$

The first step follows since

$$P(X \ge a) = \sum_{n=a}^{\infty} P(X = n) = \sum_{n=0}^{\infty} u_a(x)\, P(X = n) = E\left[u_a(X)\right]$$

for discrete random variables and

$$P(X \geq a) = \int_a^\infty f_X(x)\, dx = \int_0^\infty u_a(x) f_X(x)\, dx = E\left[u_a(X)\right]$$

for continuous random variables. The second step followed since $E[Y] \geq 0$ for non-negative random variables Y. □

Corollary 3.1.1 (Chebyshev's inequality)**.** *If X is a random variable, then for all $a > 0$,*

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}.$$

Proof. Use Markov's inequality on the non-negative random variable $(X - \mu)^2$. □

Definition 3.14 (Covariance)**.** Let X, Y be two random variables. The covariance of X and Y is defined as

$$\mathrm{Cov}[X, Y] = E\left[(X - E[X])(Y - E[Y])\right].$$

Remark. This can be simplified as

$$\mathrm{Cov}[X, Y] = E[XY] - E[X]\, E[Y].$$

Definition 3.15 (Correlation coefficient)**.** The correlation coefficient is defined as

$$\rho_{X,Y} = \frac{\mathrm{Cov}[X, Y]}{\sigma_X\, \sigma_Y}.$$

Theorem 3.2 (Weak Law of Large Numbers)**.** *Suppose that $X_1, \ldots, X_n$ are identical and independent random variables, with mean μ and variance σ^2. Define*

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Then, $E\left[\overline{X}_n\right] = \mu$, and $\mathrm{Var}[\overline{X}_n] = \sigma^2/n$. From Chebyshev,

$$\lim_{n \to \infty} P(|\overline{X}_n - \mu| \geq \epsilon) = 0.$$

Definition 3.16 (Moment generating function)**.** Let X be a random variable such that $E\left[e^{|tX|}\right]$ is finite for all t in a neighbourhood of the origin. Then, the moment generating function for X is defined in this neighbourhood as

$$M_X(t) = E\left[e^{tX}\right] = 1 + tE\left[X\right] + \frac{1}{2}t^2 E\left[X^2\right] + \dots$$

Remark. By writing the moment generating function as a Maclaurin series, we see that

$$\frac{d^n}{dt^n} M_X(t)\Big|_{t=0} = E\left[X^n\right].$$

Also note that for independent random variables,

$$M_{X+Y} = M_X \cdot M_Y.$$

Example. Consider a Bernoulli random variable X. Its moment generating function is given by

$$M_X(t) = pe^t + q.$$

If Y is a binomial random variable, then its moment generating function is simply

$$M_Y(t) = (pe^t + q)^n.$$

If Z is a Poisson random variable, then

$$M_Z(t) = e^{\lambda(e^t - 1)}.$$

If W is a unit normal random variable, then

$$M_W(t) = e^{t^2/2}.$$

Theorem 3.3. *A moment generating function determines its corresponding random variable completely. They are related via the Laplace transform and its inverse.*

$$M_X(t) = \mathcal{L}\{f_X\}(-t).$$

Example. Suppose that $X_i \sim \text{Poission}(\lambda_i)$, and let $X = X_1 + \cdots + X_n$. Then, the moment generating function of X must be the product of the moment generating functions of X_i, so

$$M_X(t) = \prod e^{\lambda_i(e^t - 1)} = e^{\lambda(e^t - 1)}$$

where $\lambda = \lambda_1 + \cdots + \lambda_n$. This is the moment generating function of a Poisson random variable as well, so $X \sim \text{Poisson}(\lambda)$.

Example. Let $X_i \sim \text{Normal}(\mu_i, \sigma_i^2)$, and let $X = X_1 + \cdots + X_n$. Note that

$$M_{X_i}(t) = e^{t\mu_i + t^2\sigma_i^2/2}.$$

Hence, the moment generating function of the sum is simply

$$M_X(t) = e^{t\mu + t^2\sigma^2/2},$$

where $\mu = \mu_1 + \cdots + \mu_n$ and $\sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2$. Thus, $X \sim \text{Normal}(\mu, \sigma^2)$.

Theorem 3.4 (Central Limit Theorem)**.** *Let $X_1, \ldots, X_n$ be identical and independent random variables with mean μ and variance σ^2. Then, the distribution of $(\overline{X} - \mu)/(\sigma/\sqrt{n})$ tends to the standard normal distribution as $n \to \infty$.*

$$\lim_{n\to\infty} P\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq a\right) = \Phi(a).$$

Proof. Note that

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i,$$

where $Y_i = (X_i - \mu)/\sigma$. Let $M(t)$ be the moment generating function of Y_i, so the corresponding moment generating function for $Y_i/\sqrt{n}$ is given as $M(t/\sqrt{n})$. Since Y_i are identical and independent, this means that the moment generating function of the sum must be $(M(t/\sqrt{n}))^n$. Setting $L(t) = \log M(t)$, we have $L(0) = 0$. Also, $E[Y_i] = 0$ and $\text{Var}[Y_i] = 1$, so $L'(0) = 0$ and $L''(0) = 1$. We wish to show that the moment generating function $((M(t/\sqrt{n})))^n$ tends to $e^{t^2/2}$ as $n \to \infty$, i.e.

$$\lim_{n\to\infty} nL(t/\sqrt{n}) = t^2/2.$$

Setting $x = 1/\sqrt{n}$ and applying L'Hôpital's rule twice,

$$\lim_{n\to\infty} nL(t/\sqrt{n}) = \lim_{x\to 0} \frac{L(tx)}{x^2} = \lim_{x\to 0} \frac{tL'(tx)}{2x} = \lim_{x\to 0} \frac{t^2 L''(tx)}{2} = t^2/2. \qquad \square$$