CHAPTER 4. RANDOM VARIABLES

When the possible outcomes of an experiment (or trial) can be given in numerical terms, then we have a random variable in hand. When an experiment is performed, the outcome of the random variable is called a "realization." A random variable can be either discreet, or continuous. A random variable is governed by its probability laws.

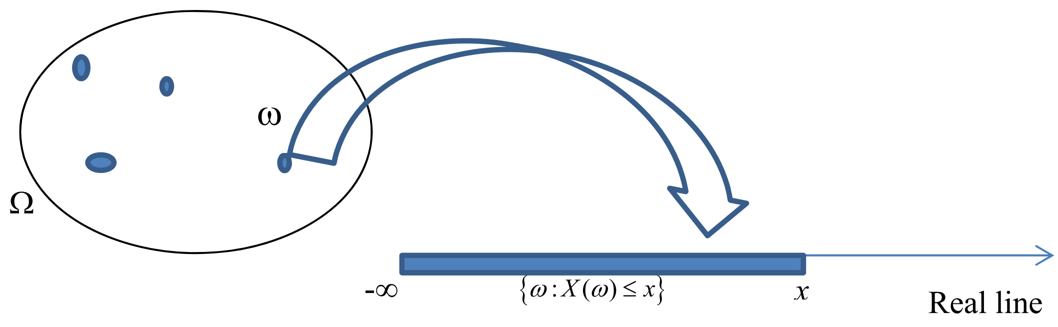
If a quantity varies randomly with time, we model it as a stochastic process. A stochastic process can be viewed as a family of random variables.

If a quantity varies randomly in space, we model it as a random field, which is the generalization of a stochastic process in two or more dimensions.

Formally, a measurable function[5] defined on a sample space is called a random variable (Feller, vol 1, p. 212). That is, X is a random variable if $X = X(\omega)$ is a function defined on the sample space Ω, and for every real x, the set

$$\{\omega : X(\omega) \le x\}$$

is an event in Ω. Thus we confine ourselves to σ-algebra of events of the type $X \le x$. Unless explicitly required, we suppress the argument ω when referring to a random variable in the rest of this text.



Random variables (RVs) are classified as discrete or continuous depending on whether their sample space is countable or uncountable, respectively. Discrete random variables typically arise from counting processes such that their range is the set of natural numbers, although a discrete RV can assume any set of discrete values on the real line, not necessarily integers. Continuous RVs on the other hand typically arise from measurement processes and their range is continous intervals (possibly infinite) on the real line. It may be useful to define "mixed" random variables if they exhibit properties of discrete and continuous RVs in different ranges.

4.1 Probability laws for RVs

A random variable is governed by its probability laws. The probability law of a RV can be described by any of the four equivalent ways:

---

[5] Measurable functions have been defined in Section 2.6

1. CDF (cumulative distribution function)
2. PDF/PMF (probability density function for continuous rv's, probability mass function for discrete rv's)
3. CF (characteristic function)
4. MGF (moment generating function)

CF and MGF are introduced after the disucssion on moments of random variables in Section 4.2.

4.1.1 CDF – cumulative distribution function

The cumulative distribution function of the random variable X is defined as:

$$F_X(x) = P[X \leq x] \tag{4.1}$$

It starts from 0, ends at 1, and is a non-decreasing function of x. It is piecewise continuous for discrete RVs, and continuous for continuous RVs.

Properties of CDF:

$$\begin{aligned} &F_X(-\infty) = 0 \\ &F_X(\infty) = 1 \\ &F_X(x) \text{ is a non-decreasing function of } x \end{aligned} \tag{4.2}$$

Thus, the probability of finding the random variable X in the semi-open interval (a,b] is:

$$P[a < X \leq b] = F_X(b) - F_X(a) \tag{4.3}$$

4.1.2 Probability density and mass functions

4.1.2.1 <u>PDF</u>

The probability density function (of continuous random variables) is defined as the derivative of the CDF:

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{4.4}$$

so that:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \tag{4.5}$$

and,

$$\int_a^b f_X(x)dx = F_X(b) - F_X(a) \tag{4.6}$$

PDF may also be interpreted as giving rise to the small probability of observing the random variable *X* around the point *x*:

$$f_X(x)dx \approx P[x + dx \le X < x] \tag{4.7}$$

4.1.2.2 PMF

The probability mass function (of discrete random variables) is defined as the probability that the random variable assumes a particular value:

Probability mass function (pmf):

$$p_X(x_i) = P[X = x_i]$$

Cumulative distribution function (cdf):

$$F_X(x_i) = P[X \le x_i]$$

so that the PMF can be derived from the CDF as:

$$p_X(x_i) = F_X(x_i) - F_X(x_i - \varepsilon) \text{ where } \varepsilon < \min(|x_i - x_j|), i \ne j \tag{4.8}$$

If the sample space of the discrete rv X is $\{x_1, x_2, ...\}$ so that $x_i \le x_{i+1}$, then the PMF may be given as the height of the step in the CDF curve:

$$p_X(x_i) = F_X(x_i) - F_X(x_{i-1}) \tag{4.9}$$

4.1.2.3 Use of delta functions to describe pmfs as pdfs

The delta function is defined as

$$\delta(x) = dU(x)/dx \tag{4.10}$$

where *U* is the unit step function:

$$U = \begin{cases} 0 \text{ if } x < 0 \\ 1 \text{ if } x \ge 0 \end{cases} \tag{4.11}$$

The delta function is symmetric: $\delta(x) = \delta(-x)$. Integrating the delta function from a to b, if $a < 0 \le b$, gives unity. More generally:

$$g(0)=\int_a^b \delta(x)g(x)dx \tag{4.12}$$

If X is a discrete RV, and $p_i = p_X(x_i)$, we can write an equivalent pdf as

$$f_X(x)=\sum_i p_i\delta(x-x_i) \tag{4.13}$$

But, in the following, we will still write formulas and expressions separately for discrete and continuous RVs.

4.2 Expectation

The expectation of any function $g(X)$ of the random variable X is defined as:

$$\begin{aligned} E[g(X)] &= \int_{-\infty}^{\infty} g(x)f_X(x)dx \text{ if X continuous} \\ &= \sum_{all\,x_i} g(x_i)p_X(x_i) \text{ if X discrete} \end{aligned} \tag{4.14}$$

The expectation of a constant is the identity operator:

$$E(c)=c \text{ where } c \text{ is a constant} \tag{4.15}$$

Expectation is a linear operator:

$$E(aX+b)= a\,E(X)+ b \tag{4.16}$$

and if $Y = g_1(X) + g_2(X) + \ldots$, then

$$E(Y)=E(g_1(X)) + E(g_2(X)) + \ldots \tag{4.17}$$

Thus the mean of X is its expectation:

$$\mu=E(X)=\begin{cases} \int_{-\infty}^{\infty} x f_X(x)dx, & \text{continuous RV} \\ \sum_{all\,x_i} x_i p_X(x_i), & \text{discrete RV} \end{cases} \tag{9.18}$$

and its variance is the expecation of its squared deviation from the mean:

$$\sigma^2 = E\left[(X-\mu)^2\right] = \begin{cases} \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx, & \text{continuous RV} \\ \sum_{all\, x_i} (x_i-\mu)^2 p_X(x_i), & \text{discrete RV} \end{cases} \tag{9.19}$$

Higher order raw or central moments can be defined by choosing $g(X) = X^n$ or $(X-\mu)^n$ above as appropriate. The characteristic function and the moment generating function can both be described as expectations of appropriate functions of *X*.

4.3 Moments

The k^{th} central moment of X:

$$\mu_{c,k} = E\left[(X-\mu)^k\right] \tag{4.20}$$

Hence, by definition, $\mu_{c,0} = 1$ and $\mu_{c,1} = 0$. The variance of X is the second central moment:

$$\text{Variance, } \sigma^2 = \mu_{c,2} = \begin{cases} \sum_{i=1}^{n} p_i (x_i-\mu)^2 & \text{for discrete RV} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx & \text{for continuous RV} \end{cases} \tag{4.21}$$

The k^{th} raw moment of X: $\mu_{0,k} = E\left[(X)^k\right]$. Hence, the mean of X is simply the first raw moment:

$$\mu = \mu_{0,1} = \begin{cases} \sum_{i=1}^{n} p_i x_i & \text{for discrete RV} \\ \int_{-\infty}^{\infty} x\, f_X(x)dx & \text{for continuous RV} \end{cases}$$

General formula for central moments (by simply expanding the binomial series[6]):

$$\mu_{c,n} = \sum_{k=0}^{n} \binom{n}{k} \mu_{0,k} (-\mu)^{n-k} \tag{4.22}$$

Similarly, the general formula for raw moments:

$$\mu_{0,n} = \sum_{k=0}^{n} \binom{n}{k} \mu_{c,k} (\mu_X)^{n-k} \tag{4.23}$$

[6] If *n* is an integer, the binomial series is given by: $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$

In particular,

$$\begin{aligned} \mu_{c,3} &= \mu_{0,3} - 3\mu_{0,2}\mu + 2\mu^3 \\ \mu_{0,3} &= \mu^3 + 3\mu\sigma^2 + \mu_{c,3} \end{aligned} \quad \text{, etc.}$$

4.4 Characteristic function

PDF/PMF and the Characteristic Function, M_X form a Fourier Transform (FT) pair. In the continuous case, the characteristic function is the Fourier transform of the pdf:

$$M_X(\theta) = E[\exp(i\theta X] = \int_{-\infty}^{\infty} e^{i\theta x} f_X(x)dx \tag{4.24}$$

such that the inverse transform is:

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} M_X(\theta)e^{-i\theta x}d\theta \tag{4.25}$$

If f_X is discontinous at some x_1, this equals $\frac{1}{2}[f_X(x_1+) + f_X(x_1-)]$. In Eq (4.24) E is the expectation operator discussed in Section 4.2. The requirement is that f_X is absolutely integrable, i.e., $\int_{-\infty}^{\infty} |f_X(x)|\,dx < \infty$, which is no problem since f_X is a pdf to begin with.

In the discrete case, the FT pair is given by:

$$\begin{aligned} M_X(\theta) &= \sum_{\text{all}\,x_j} p_X(x_j)e^{i\theta x_j} \\ p_X(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} M_X(\theta)e^{-i\theta x}d\theta \end{aligned} \tag{4.26}$$

Proof:

$$\begin{aligned} &\frac{1}{2\pi}\int_{-\infty}^{\infty} M_X(\theta)e^{-i\theta x}d\theta \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\sum_{\text{all}\,x_j} p_X(x_j)e^{i\theta x_j}e^{-i\theta x}d\theta \\ &= \frac{1}{2\pi}\sum_{\text{all}\,x_j} p_X(x_j)\int_{-\infty}^{\infty} e^{i(x_j - x)\theta}d\theta \\ &= \frac{1}{2\pi}\sum_{\text{all}\,x_j} p_X(x_j)2\pi\delta(x - x_j) \\ &= p_X(x) \end{aligned} \tag{4.27}$$

The characteristic function of a RV uniquely determines its probability distribution.

4.5 Moment generating function

The moment generating function (MGF) of a distribution is the expectation:

$$G_X(s) = E\left[e^{sX}\right] = \int_{-\infty}^{\infty} e^{sX} f_X(x)dx \tag{4.28}$$

It exists if the integral is finite for all s in some interval I that contains 0 in its interior (Resnick, p. 294). If it exists, the MGF uniquely determines the distribution of X. In comparison, the CF of a RV *always* exists.

MGF is infinitely differentiable and the n^{th} derivative,

$$\frac{d^n}{ds^n} G_X(s) = \int_{-\infty}^{\infty} x^n e^{sx} f_X(x)dx \tag{4.29}$$

evaluated at s = 0, gives the n^{th} raw moment of X:

$$\frac{d^n}{ds^n} G_X(s)|_{s=0} = E\left[X^n\right] \tag{4.30}$$

If instead we perform successive derivatives on $\ln G_X(s)$ and evaluate them at s = 0, we get the central moments of X:

$$\frac{d}{ds} \ln G_X(s) = \mu_X \quad \text{at } s = 0$$

$$\frac{d^2}{ds^2} \ln G_X(s) = \sigma_X^2 \text{ at } s = 0$$

$$\frac{d^3}{ds^3} \ln G_X(s) = E\left[\left(X - \mu_X\right)^3\right] \; at \;\; s = o$$

etc.

where the identity $G_X(0) = 1$ has been used.

4.5.1.1 MGF and density function

Prove that if one did not know the functional form of a given distribution, one would need moments of all orders to completely specify the distribution.

Proof: Expand the MGF in its Taylor series around the origin:

$$G_X(s) = \sum_{n=0}^{\infty} \frac{s^n}{n!} \frac{d^n}{ds^n} G_X(s)|_{s=0} = \sum_{n=0}^{\infty} \frac{s^n}{n!} E\left[X^n\right] \tag{4.31}$$

where Eq 4.20 has been used. This is valid if all moments are finite and the series converges absolutely near s = 0. Since f_X can be determined in terms of G_X, Eq 4.21

shows that all moments of X are required for the complete specification of f_X under the conditions stated above.

4.5.1.2 Cramer's Theorem:

Let X_1 and X_2 be r.v.s with PDFs f_{X_1} and f_{X_2}
and MGFs G_{X_1} and G_{X_2}

If $G_{X_1} = G_{X_2}$

then $f_{X_1} = f_{X_2}$

4.5.2 Moments from Characteristic function

The mean can be obtained as:

$$E(X) = \mu_X = \frac{1}{i}\left(\frac{dM_X}{d\theta}\right)_{\theta=0} \tag{4.32}$$

which can also be obtained using log transformation:

$$E(X) = \frac{1}{i}\frac{d}{d\theta}\left(\ln M_X(\theta)\right)_{\theta=0} \tag{4.33}$$

Higher moments are given by:

$$E\left(X^n\right) = \frac{1}{i^n}\left(\frac{d^n M_X}{d\theta^n}\right)_{\theta=0} \tag{4.34}$$

and in particular, the variance can be expressed through the log transformation as:

$$\mathrm{var}(X) = \frac{1}{i^2}\frac{d^2}{d\theta^2}\left(\ln M_X(0)\right) \tag{4.35}$$

4.6 Basic properties of random variables

4.6.1 Markov's inequality

If $f_X(x) = 0$ for $x < 0$,

$$P\left[X \ge \varepsilon\mu\right] \le \frac{1}{\varepsilon},\quad \varepsilon > 0 \tag{4.36}$$

4.6.2 Chebyshev's inequality

$$P\left[\left|X - \mu\right| \ge \varepsilon\sigma\right] \le \frac{1}{\varepsilon^2},\quad \varepsilon > 0 \tag{4.37}$$

4.6.2.1 Example involving a discrete RV

Take the Bernoulli RV $X(p)$ for which $P[X=0]=1-p,\ P[X=1]=p$ so that its mean is p and variance is pq. Chebyshev's inequality states:

$$P\left[\,|X-p|>\varepsilon\sqrt{pq}\,\right]\le\frac{1}{\varepsilon^2}\qquad \text{for any}\quad \varepsilon>0 \tag{4.38}$$

We prove it by considering three different ranges for *X*.

Case (1): $X > p$

The LHS of (4.38) =

$$P\left[X>p+\varepsilon\sqrt{pq}\right]=\begin{cases}p \text{ if } p+\varepsilon\sqrt{pq}<1 & \text{Case 1a}\\ 0 \text{ if } p+\varepsilon\sqrt{pq}\ge 1 & \text{Case 1b}\end{cases}$$

Case (1a)

LHS $=p$ and we need to prove $p\le\frac{1}{\varepsilon^2}$

It is given that, $p+\varepsilon\sqrt{pq}<1$

$\Rightarrow \varepsilon<\frac{1-p}{\sqrt{pq}}=\frac{q}{\sqrt{pq}}=\sqrt{\frac{q}{p}}$

$\Rightarrow \varepsilon^2<\frac{q}{p}$

$\Rightarrow p<\frac{q}{\varepsilon^2}$

Since $q<1$, we have $p<\frac{1}{\varepsilon^2}$ ... proved

Case (1b)

LHS $=0$ and we need to prove $0\le\frac{1}{\varepsilon^2}$

It is given that, $p+\varepsilon\sqrt{pq}\ge 1$

$\Rightarrow \varepsilon\ge\frac{1-p}{\sqrt{pq}}=\frac{q}{\sqrt{pq}}=\sqrt{\frac{q}{p}}$

$\Rightarrow \varepsilon^2\ge\frac{q}{p}$

Since ε^2 is a positive quantity, we have $0<\frac{1}{\varepsilon^2}$ ... proved

Case (2): $0 \le X < p$

LHS =

$$P\left[p - X > \varepsilon\sqrt{pq}\right] = P\left[X < p - \varepsilon\sqrt{pq}\right]$$

$$= \begin{cases} 0 \text{ if } \ p < \varepsilon\sqrt{pq} & \text{Case } 2a \\ q \text{ if } \ p \ge \varepsilon\sqrt{pq} & \text{Case } 2b \end{cases}$$

Case (2a)

LHS $= 0$ and we need to prove $0 \le \dfrac{1}{\varepsilon^2}$

It is given that $p < \varepsilon\sqrt{pq}$

$\Rightarrow \dfrac{p}{q} \le \varepsilon^2$

Since ε^2 is a positive quantity, we have $0 < \dfrac{1}{\varepsilon^2}$ ... proved

Case (2b)

LHS $= q$ and we need to prove $q \le \dfrac{1}{\varepsilon^2}$

It is given that $p \ge \varepsilon\sqrt{pq}$

$\Rightarrow \dfrac{p}{q} \ge \varepsilon^2 \Rightarrow \dfrac{p}{\varepsilon^2} \ge q$

Since $p \le 1$, we have $q \le \dfrac{1}{\varepsilon^2}$ ... proved

Case (3)

$X < 0$

$LHS \ = P\left[X < p - \varepsilon\sqrt{pq}\right] = 0 \le \dfrac{1}{\varepsilon^2}$

$\Rightarrow p < \varepsilon\sqrt{pq}$

$\Rightarrow \dfrac{p}{q} < \varepsilon^2$

$i.e. \ \dfrac{1}{\varepsilon^2} \ge 0 \ \ if \ \ \varepsilon^2 > p/q$

Proved.

4.6.3 Bienyame Inequality

$$P\left[|X - a|^n \ge \varepsilon^n\right] \le \frac{E[|X - a|^n]}{\varepsilon^n}, \ \ \varepsilon > 0 \tag{4.39}$$

n=2 reduces to Chebyshev's inequality.

4.6.4 Lyapunov Inequality

Let $\beta_k = E[|X|^k] < \infty$ where k => 1 is any integer, then

$$\beta_{k-1}^{1/(k-1)} \le \beta_k^{1/k} \tag{4.40}$$

Proof

Let $T = a|X|^{\frac{k+1}{2}} + |X|^{\frac{k-1}{2}}$

$T^2 = a^2|X|^{k+1} + 2a|X|^k + |X|^{k-1}$

$E\left(T^2\right) = a^2\beta_{k+1} + 2a\beta_k + \beta_{k-1} \ge 0$

Dividing by β_{k+1} we get,

$$a^2 + 2a\frac{\beta_k}{\beta_{k+1}} + \frac{\beta_k^2}{\beta_{k+1}^2} + \frac{\beta_{k-1}}{\beta_{k+1}} - \frac{\beta_k^2}{\beta_{k+1}^2} \ge 0$$

Now choose, $a = -\dfrac{\beta_k}{\beta_{k+1}}$ which yields

$$\frac{\beta_{k-1}}{\beta_{k+1}} - \frac{\beta_k^2}{\beta_{k+1}^2} \ge 0 \Rightarrow \boxed{\beta_{k-1}\beta_{k+1} \ge \beta_k^2} \tag{4.41}$$

Set *k*=1,2,3 etc.:

$k = 1 \Rightarrow \beta_0\beta_2 \ge \beta_1^2, \quad \beta_0 = 1 \qquad \beta_1 \le \beta_2^{1/2}$

$k = 2 \Rightarrow \beta_1\beta_3 \ge \beta_2^2 \Rightarrow \beta_2^{1/2}\beta_3 \ge \beta_2^2 \Rightarrow \beta_2^{3/2} \le \beta_3 \Rightarrow \beta_2^{1/2} \le \beta_3^{1/3}$

$k = 3 \Rightarrow \beta_2\beta_4 \ge \beta_3^2 \Rightarrow \beta_3^{2/3}\beta_4 \ge \beta_3^2 \Rightarrow \beta_3^{4/3} \le \beta_4 \Rightarrow \beta_3^{1/3} \le \beta_4^{1/4}$

We see that proposition 4.40 is true for *k* = 1,2,3. Let us claim it is true for some *k* - 1: i.e.,

$$\beta_{k-2}^{1/k-2} \le \beta_{k-1}^{1/k-1} \tag{4.42}$$

is true. We can rewrite 4.41 for $k-1$ as:

$$\beta_{k-2}\beta_k \geq \beta_{k-1}^2$$
$$\text{or, } \beta_{k-1}^{k-2/k-1}\beta_k \geq \beta_{k-1}^2$$
$$\text{or, } \beta_k \geq \left(\beta_{k-1}\right)^{2-\left(\frac{k-2}{k-1}\right)} = \beta_{k-1}^{\frac{2k-2-k+2}{k-1}} = \beta_{k-1}^{k/k-1}$$
$$\Rightarrow \boxed{\beta_k^{1/k} \geq \beta_{k-1}^{1/k-1}} \tag{4.43}$$

Thus we prove (4.40) by induction: if (4.40) is true for $k-1$, then it must be true for k. Since we have proved (4.40) is true for k =3, it must be true for 4, 5, ….

Proved.

Since $\beta_n^{1/n} \geq \beta_{n-1}^{1/n-1} \geq \beta_{n-2}^{1/n-2} \geq \beta_m^{1/m} \geq \ldots \geq \beta_3^{1/3} \geq \beta_2^{1/2} \geq \beta_1 \geq 1$ and for $n \geq m$, an equivalent way of writing Lyapunov inequality is:

$$\beta_n^{1/n} \geq \beta_m^{1/m} \Rightarrow \boxed{\beta_n^m \geq \beta_m^n} \tag{4.44}$$