Chapter 11

General Random Variables

11.1 Law of a Random Variable

Thus far we have considered only random variables whose domain and range are discrete. We now consider a general random variable $X : \Omega \to I\!R$ defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Recall that:

- $\mathcal{F}$ is a σ-algebra of subsets of Ω.
- P is a probability measure on $\mathcal{F}$, i.e., $I\!P(A)$ is defined for every $A \in \mathcal{F}$.

A function $X : \Omega \to I\!R$ is a random variable if and only if for every $B \in \mathcal{B}(I\!R)$ (the σ-algebra of Borel subsets of R), the set

$$\{X \in B\} \stackrel{\triangle}{=} X^{-1}(B) \stackrel{\triangle}{=} \{\omega; X(\omega) \in B\} \in \mathcal{F},$$

i.e., $X : \Omega \to I\!R$ is a random variable if and only if X^{-1} is a function from $\mathcal{B}(I\!R)$ to $\mathcal{F}$(See Fig. 11.1)

Thus any random variable X induces a measure μ_X on the measurable space $(I\!R, \mathcal{B}(I\!R))$ defined by

$$\mu_X(B) = I\!P\left(X^{-1}(B)\right) \quad \forall B \in \mathcal{B}(I\!R),$$

where the probabiliy on the right is defined since $X^{-1}(B) \in \mathcal{F}$. μ_X is often called the *Law of* X – in Williams' book this is denoted by $\mathcal{L}_X$.

11.2 Density of a Random Variable

The *density of* X (if it exists) is a function $f_X : I\!R \to [0, \infty)$ such that

$$\mu_X(B) = \int_B f_X(x)\, dx \quad \forall B \in \mathcal{B}(I\!R).$$

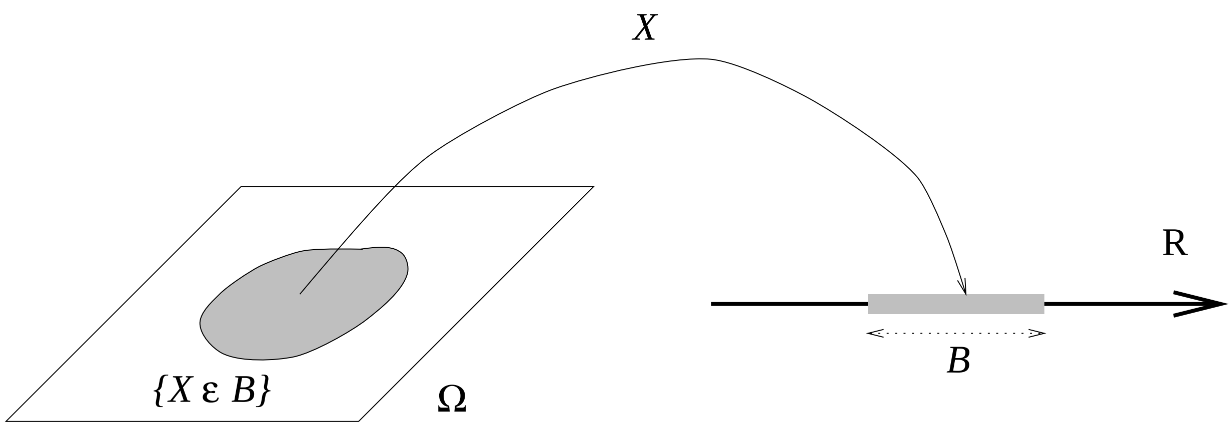

Figure 11.1: *Illustrating a real-valued random variable X.*

We then write

$$d\mu_X(x) = f_X(x)dx,$$

where the integral is with respect to the Lebesgue measure on R. f_X is the Radon-Nikodym derivative of μ_X with respect to the Lebesgue measure. Thus X has a density if and only if μ_X is absolutely continuous with respect to Lebesgue measure, which means that whenever $B \in \mathcal{B}(\mathbb{R})$ has Lebesgue measure zero, then

$$\mathbb{P}\{X \in B\} = 0.$$

11.3 Expectation

Theorem 3.32 (Expectation of a function of X) *Let $h : \mathbb{R} \to \mathbb{R}$ be given. Then*

$$\begin{aligned} \mathbb{E}h(X) &\triangleq \int_\Omega h(X(\omega))\, d\mathbb{P}(\omega) \\ &= \int_{\mathbb{R}} h(x)\, d\mu_X(x) \\ &= \int_{\mathbb{R}} h(x) f_X(x)\, dx. \end{aligned}$$

Proof: (Sketch). If $h(x) = \mathbf{1}_B(x)$ for some $B \subset \mathbb{R}$, then these equations are

$$\begin{aligned} \mathbb{E}\mathbf{1}_B(X) &\triangleq P\{X \in B\} \\ &= \mu_X(B) \\ &= \int_B f_X(x)\, dx, \end{aligned}$$

which are true by definition. Now use the "standard machine" to get the equations for general h.
■

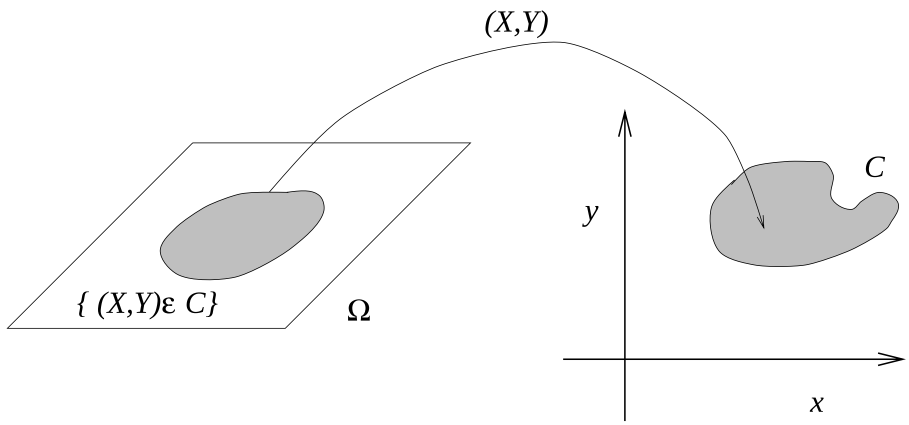

Figure 11.2: *Two real-valued random variables* X, Y.

11.4 Two random variables

Let X, Y be two random variables $\Omega \to \mathbb{R}$ defined on the space $(\Omega, \mathcal{F}, \mathbb{P})$. Then X, Y induce a measure on $\mathcal{B}(\mathbb{R}^2)$ (see Fig. 11.2) called the *joint law of* (X, Y), defined by

$$\mu_{X,Y}(C) \stackrel{\triangle}{=} \mathbb{P}\{(X,Y) \in C\} \quad \forall C \in \mathcal{B}(\mathbb{R}^2).$$

The *joint density of* (X, Y) is a function

$$f_{X,Y} : \mathbb{R}^2 \to [0, \infty)$$

that satisfies

$$\mu_{X,Y}(C) = \iint_C f_{X,Y}(x,y)\, dxdy \quad \forall C \in \mathcal{B}(\mathbb{R}^2).$$

$f_{X,Y}$ is the Radon-Nikodym derivative of $\mu_{X,Y}$ with respect to the Lebesgue measure (area) on $\mathbb{R}^2$.

We compute the expectation of a function of X, Y in a manner analogous to the univariate case:

$$\begin{aligned} \mathbb{E}k(X,Y) &\stackrel{\triangle}{=} \int_\Omega k(X(\omega), Y(\omega))\, d\mathbb{P}(\omega) \\ &= \iint_{\mathbb{R}^2} k(x,y)\, d\mu_{X,Y}(x,y) \\ &= \iint_{\mathbb{R}^2} k(x,y) f_{X,Y}(x,y)\, dxdy \end{aligned}$$

11.5 Marginal Density

Suppose (X,Y) has joint density $f_{X,Y}$. Let $B \subset \mathbb{R}$ be given. Then

$$\begin{aligned}\mu_Y(B) &= \mathbb{P}\{Y \in B\} \\ &= \mathbb{P}\{(X,Y) \in \mathbb{R} \times B\} \\ &= \mu_{X,Y}(\mathbb{R} \times B) \\ &= \int_B \int_{\mathbb{R}} f_{X,Y}(x,y)\, dxdy \\ &= \int_B f_Y(y)\, dy,\end{aligned}$$

where

$$f_Y(y) \stackrel{\triangle}{=} \int_{\mathbb{R}} f_{X,Y}(x,y)\, dx.$$

Therefore, $f_Y(y)$ is the (marginal) density for Y.

11.6 Conditional Expectation

Suppose (X,Y) has joint density $f_{X,Y}$. Let $h : \mathbb{R} \to \mathbb{R}$ be given. Recall that $\mathbb{E}[h(X)|Y] \stackrel{\triangle}{=} \mathbb{E}[h(X)|\sigma(Y)]$ depends on ω through Y, i.e., there is a function $g(y)$ (g depending on h) such that

$$\mathbb{E}[h(X)|Y](\omega) = g(Y(\omega)).$$

How do we determine g?

We can characterize g using *partial averaging:* Recall that $A \in \sigma(Y) \Longleftrightarrow A = \{Y \in B\}$ for some $B \in \mathcal{B}(\mathbb{R})$. Then the following are equivalent characterizations of g:

$$\int_A g(Y)\, d\mathbb{P} = \int_A h(X)\, d\mathbb{P} \quad \forall A \in \sigma(Y), \tag{6.1}$$

$$\int_\Omega \mathbf{1}_B(Y) g(Y)\, d\mathbb{P} = \int_\Omega \mathbf{1}_B(Y) h(X)\, d\mathbb{P} \quad \forall B \in \mathcal{B}(\mathbb{R}), \tag{6.2}$$

$$\int_{\mathbb{R}} \mathbf{1}_B(y) g(y) \mu_Y(dy) = \iint_{\mathbb{R}^2} \mathbf{1}_B(y) h(x)\, d\mu_{X,Y}(x,y) \quad \forall B \in \mathcal{B}(\mathbb{R}), \tag{6.3}$$

$$\int_B g(y) f_Y(y)\, dy = \int_B \int_{\mathbb{R}} h(x) f_{X,Y}(x,y)\, dxdy \quad \forall B \in \mathcal{B}(\mathbb{R}). \tag{6.4}$$

11.7 Conditional Density

A function $f_{X|Y}(x|y) : \mathbb{R}^2 \to [0,\infty)$ is called a *conditional density* for X given Y provided that for any function $h : \mathbb{R} \to \mathbb{R}$:

$$g(y) = \int_{\mathbb{R}} h(x) f_{X|Y}(x|y)\, dx. \tag{7.1}$$

(Here g is the function satisfying

$$\mathbb{E}[h(X)|Y] = g(Y),$$

and g depends on h, but $f_{X|Y}$ does not.)

Theorem 7.33 *If (X,Y) has a joint density $f_{X,Y}$, then*

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}. \tag{7.2}$$

Proof: Just verify that g defined by (7.1) satisfies (6.4): For $B \in \mathcal{B}(\mathbb{R})$,

$$\int_B \underbrace{\int_{\mathbb{R}} h(x) f_{X|Y}(x|y)\, dx}_{g(y)}\, f_Y(y)\, dy \;=\; \int_B \int_{\mathbb{R}} h(x) f_{X,Y}(x,y)\, dx dy.$$

Notation 11.1 Let g be the function satisfying

$$\mathbb{E}[h(X)|Y] = g(Y).$$

The function g is often written as

$$g(y) = \mathbb{E}[h(X)|Y = y],$$

and (7.1) becomes

$$\mathbb{E}[h(X)|Y = y] = \int_{\mathbb{R}} h(x) f_{X|Y}(x|y)\, dx.$$

In conclusion, to determine $\mathbb{E}[h(X)|Y]$ (a function of ω), first compute

$$g(y) = \int_{\mathbb{R}} h(x) f_{X|Y}(x|y)\, dx,$$

and then replace the dummy variable y by the random variable Y:

$$\mathbb{E}[h(X)|Y](\omega) = g(Y(\omega)).$$

Example 11.1 (Jointly normal random variables) Given parameters: $\sigma_1 > 0, \sigma_2 > 0, -1 < \rho < 1$. Let (X,Y) have the joint density

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{x^2}{\sigma_1^2} - 2\rho \frac{x}{\sigma_1}\frac{y}{\sigma_2} + \frac{y^2}{\sigma_2^2} \right] \right\}.$$

The exponent is

$$\begin{aligned} & -\frac{1}{2(1-\rho^2)}\left[\frac{x^2}{\sigma_1^2}-2\rho\frac{x}{\sigma_1}\frac{y}{\sigma_2}+\frac{y^2}{\sigma_2^2}\right] \\ &= -\frac{1}{2(1-\rho^2)}\left[\left(\frac{x}{\sigma_1}-\rho\frac{y}{\sigma_2}\right)^2+\frac{y^2}{\sigma_2^2}(1-\rho^2)\right] \\ &= -\frac{1}{2(1-\rho^2)}\frac{1}{\sigma_1^2}\left(x-\frac{\rho\sigma_1}{\sigma_2}y\right)^2-\frac{1}{2}\frac{y^2}{\sigma_2^2}. \end{aligned}$$

We can compute the *Marginal density of* Y as follows

$$\begin{aligned} f_Y(y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{\infty} e^{-\frac{1}{2(1-\rho^2)\sigma_1^2}\left(x-\frac{\rho\sigma_1}{\sigma_2}y\right)^2}\,dx.e^{-\frac{1}{2}\frac{y^2}{\sigma_2^2}} \\ &= \frac{1}{2\pi\sigma_2}\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}}\,du.e^{-\frac{1}{2}\frac{y^2}{\sigma_2^2}} \\ &\qquad \text{using the substitution } u=\frac{1}{\sqrt{1-\rho^2}\sigma_1}\left(x-\frac{\rho\sigma_1}{\sigma_2}y\right),\ du=\frac{dx}{\sqrt{1-\rho^2}\sigma_1} \\ &= \frac{1}{\sqrt{2\pi}\,\sigma_2}e^{-\frac{1}{2}\frac{y^2}{\sigma_2^2}}. \end{aligned}$$

Thus Y is normal with mean 0 and variance σ_2^2.

Conditional density. From the expressions

$$f_{X,Y}(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\frac{1}{\sigma_1^2}\left(x-\frac{\rho\sigma_1}{\sigma_2}y\right)^2}e^{-\frac{1}{2}\frac{y^2}{\sigma_2^2}},$$

$$f_Y(y)=\frac{1}{\sqrt{2\pi}\,\sigma_2}e^{-\frac{1}{2}\frac{y^2}{\sigma_2^2}},$$

we have

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \frac{1}{\sqrt{2\pi}\,\sigma_1}\frac{1}{\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\frac{1}{\sigma_1^2}\left(x-\frac{\rho\sigma_1}{\sigma_2}y\right)^2}. \end{aligned}$$

In the x-variable, $f_{X|Y}(x|y)$ is a normal density with mean $\frac{\rho\sigma_1}{\sigma_2}y$ and variance $(1-\rho^2)\sigma_1^2$. Therefore,

$$\mathbb{E}[X|Y=y]=\int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx = \frac{\rho\sigma_1}{\sigma_2}y;$$

$$\begin{aligned} & \mathbb{E}\left[\left(X-\frac{\rho\sigma_1}{\sigma_2}y\right)^2 \middle| Y=y\right] \\ &= \int_{-\infty}^{\infty}\left(x-\frac{\rho\sigma_1}{\sigma_2}y\right)^2 f_{X|Y}(x|y)\,dx \\ &= (1-\rho^2)\sigma_1^2. \end{aligned}$$

From the above two formulas we have the formulas

$$\mathbb{E}[X|Y] = \frac{\rho\sigma_1}{\sigma_2}Y, \tag{7.3}$$

$$\mathbb{E}\left[\left(X - \frac{\rho\sigma_1}{\sigma_2}Y\right)^2 \Big| Y\right] = (1-\rho^2)\sigma_1^2. \tag{7.4}$$

Taking expectations in (7.3) and (7.4) yields

$$\mathbb{E}X = \frac{\rho\sigma_1}{\sigma_2}\mathbb{E}Y = 0, \tag{7.5}$$

$$\mathbb{E}\left[\left(X - \frac{\rho\sigma_1}{\sigma_2}Y\right)^2\right] = (1-\rho^2)\sigma_1^2. \tag{7.6}$$

Based on Y, the best estimator of X is $\frac{\rho\sigma_1}{\sigma_2}Y$. This estimator is unbiased (has expected error zero) and the expected square error is $(1-\rho^2)\sigma_1^2$. No other estimator based on Y can have a smaller expected square error (Homework problem 2.1). ■

11.8 Multivariate Normal Distribution

Please see Oksendal Appendix A.

Let $\mathbf{X}$ denote the column vector of random variables $(X_1, X_2, \ldots, X_n)^T$, and $\mathbf{x}$ the corresponding column vector of values $(x_1, x_2, \ldots, x_n)^T$. $\mathbf{X}$ has a multivariate normal distribution if and only if the random variables have the joint density

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\sqrt{\det \mathbf{A}}}{(\mathbf{2}\pi)^{\mathbf{n/2}}} \exp\left\{-\tfrac{1}{2}(\mathbf{X}-\boldsymbol{\mu})^{\mathbf{T}}.\mathbf{A}.(\mathbf{X}-\boldsymbol{\mu})\right\}.$$

Here,

$$\boldsymbol{\mu} \triangleq (\mu_1, \ldots, \mu_n)^T = \mathbb{E}\mathbf{X} \triangleq (\mathbb{E}X_1, \ldots, \mathbb{E}X_n)^T,$$

and A is an $n \times n$ nonsingular matrix. A^{-1} is the covariance matrix

$$A^{-1} = \mathbb{E}\left[(\mathbf{X}-\boldsymbol{\mu}).(\mathbf{X}-\boldsymbol{\mu})^T\right],$$

i.e. the (i,j)th element of A^{-1} is $\mathbb{E}(X_i-\mu_i)(X_j-\mu_j)$. The random variables in $\mathbf{X}$ are independent if and only if A^{-1} is diagonal, i.e.,

$$A^{-1} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2),$$

where $\sigma_j^2 = \mathbb{E}(X_j - \mu_j)^2$ is the variance of X_j.

11.9 Bivariate normal distribution

Take $n = 2$ in the above definitions, and let

$$\rho \triangleq \frac{\mathbb{E}(X_1 - \mu_1)(X_2 - \mu_2)}{\sigma_1 \sigma_2}.$$

Thus,

$$A^{-1} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix},$$

$$A = \begin{bmatrix} \frac{1}{\sigma_1^2(1-\rho^2)} & -\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)} \\ -\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)} & \frac{1}{\sigma_2^2(1-\rho^2)} \end{bmatrix},$$

$$\sqrt{\det A} = \frac{1}{\sigma_1\sigma_2\sqrt{1-\rho^2}},$$

and we have the formula from Example 11.1, adjusted to account for the possibly non-zero expectations:

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}.$$

11.10 MGF of jointly normal random variables

Let $\mathbf{u} = (u_1, u_2, \ldots, u_n)^T$ denote a column vector with components in $\mathbb{R}$, and let $\mathbf{X}$ have a multivariate normal distribution with covariance matrix A^{-1} and mean vector $\boldsymbol{\mu}$. Then the moment generating function is given by

$$\begin{aligned} \mathbb{E}e^{\mathbf{u}^T.\mathbf{X}} &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{\mathbf{u}^T.\mathbf{X}} f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n)\, dx_1 \ldots dx_n \\ &= \exp\left\{\tfrac{1}{2}\mathbf{u}^T A^{-1}\mathbf{u} + \mathbf{u}^T\boldsymbol{\mu}\right\}. \end{aligned}$$

If any n random variables $X_1, X_2, \ldots, X_n$ have this moment generating function, then they are jointly normal, and we can read out the means and covariances. The random variables are jointly normal *and independent* if and only if for any real column vector $\mathbf{u} = (u_1, \ldots, u_n)^T$

$$\mathbb{E}e^{\mathbf{u}^T.\mathbf{X}} \triangleq \mathbb{E}\exp\left\{\sum_{j=1}^n u_j X_j\right\} = \exp\left\{\sum_{j=1}^n [\tfrac{1}{2}\sigma_j^2 u_j^2 + u_j\mu_j]\right\}.$$