Lecture 4:
More on Continuous Random Variables and Functions of Random Variables

ELE 525: Random Processes in Information Systems

Hisashi Kobayashi

Department of Electrical Engineering
Princeton University
September 25, 2013

Textbook: Hisashi Kobayashi, Brian L. Mark and William Turin, ***Probability, Random Processes and Statistical Analysis*** (Cambridge University Press, 2012)

4.3 Joint and Conditional Probability Density Functions

- If $F_{XY}(x, y)$ is everywhere continuous and possesses a second partial derivative everywhere, we define the **joint PDF** by

$$f_{XY}(x,y) = \frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}. \tag{4.92}$$

$$F_{XY}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(u,v)\,du\,dv. \tag{4.93}$$

- The **conditional distribution function** of RV *Y,* given *X=x,* is

$$F_{Y|X}(y|x) = \frac{\int_{-\infty}^{y} f_{XY}(x,v)\,dv}{f_X(x)}, \tag{4.101}$$

$$F_{Y|X}(y|x) = \int_{-\infty}^{y} f_{Y|X}(v|x)\,dv. \tag{4.103}$$

- The **conditional expectation** of *X* given *Y* is defined by

$$\boxed{E[X|Y] \triangleq \psi(Y),} \tag{4.104}$$

where

$$\psi(y) = E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx. \tag{4.105}$$

- The **law of iterated expectations** holds:

$$E[E[X|Y]] = E[X]. \tag{4.106}$$

- The conditional expectation is the best estimate of *X* as a function of *Y* in the *minimum mean square error* (MMSE) sense (see Section 22.1.3, pp. 649-651)

4.3.1 Bivariate normal (or Gaussian) distribution

The standard bivariate normal distribution is defined by

$$\phi_\rho(u_1, u_2) \triangleq \frac{1}{2\pi\sqrt{(1-\rho^2)}} \exp\left\{-\frac{1}{2(1-\rho^2)}(u_1^2 - 2\rho u_1 u_2 + u_2^2)\right\}, \quad (4.108)$$

$$\rho \triangleq \mathrm{Cov}[U_1, U_2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u_1 u_2 \phi_\rho(u_1, u_2)\, du_1\, du_2, \quad (4.109)$$

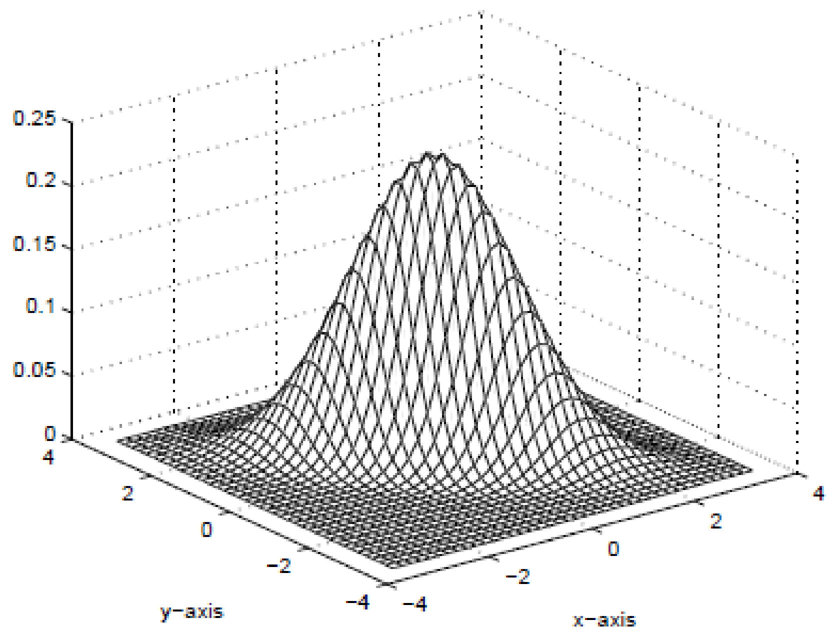


Figure 4.7 (a) The standard bivariate normal distribution $\phi_\rho(u_1, u_2)$ of (4.108) with $\rho = -0.75$.

- When ρ=0, the RVs U_1 and U_2 are said to be **uncorrelated** and

$$\phi_0(u_1, u_2) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u_1^2}{2}\right\} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u_2^2}{2}\right\} = \phi(u_1)\phi(u_2), \qquad (4.110)$$

- Thus, the bivariate normal variables are independent when they are uncorrelated. (Two uncorrelated RVs are not necessarily independent, unless they are normal RVs.)
- The conditional PDF of U_2 given U_1=u_1 can be computed as

$$\boxed{f_{U_2|U_1}(u_2|u_1) = \frac{\phi_\rho(u_1, u_2)}{\phi(u_1)} = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{(u_2-\rho u_1)^2}{2(1-\rho^2)}\right\},} \qquad (4.111)$$

which is also a normal distribution, with mean ρu_1 and variance $1-\rho^2$.

❖ Define RVs X_1 and X_2 by

$$U_1 = \frac{X_1 - \mu_1}{\sigma_1} \text{ and } U_2 = \frac{X_2 - \mu_2}{\sigma_2}. \tag{4.112}$$

Then the joint PDF of X_1 and X_2 is

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{Q(x_1,x_2)}{2}\right\}, \tag{4.113}$$

where

$$Q(x_1,x_2) = \frac{1}{1-\rho^2}\left[\left(\frac{X_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{X_1-\mu_1}{\sigma_1}\right)\left(\frac{X_2-\mu_2}{\sigma_2}\right) + \left(\frac{X_2-\mu_2}{\sigma_2}\right)^2\right]. \tag{4.114}$$

❖ Adopt a vector notation: $\boldsymbol{X} \triangleq \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ (4.115)

Then

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{2\pi|\det \boldsymbol{C}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}, \tag{4.116}$$

where ***C*** is the covariance matrix, given by

$$C = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}, \tag{4.117}$$

and

$$|\det C| = \sigma_1^2\sigma_2^2(1-\rho^2). \tag{4.118}$$

$$C^{-1} = \frac{1}{|\det C|}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} = \frac{1}{1-\rho^2}\begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}, \tag{4.119}$$

4.4 Exponential Family of Distributions

A family of PDFs (or PMFs) of the form

$$f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}) = h(\boldsymbol{x})\exp\{\boldsymbol{\eta}^\top(\boldsymbol{\theta})\boldsymbol{T}(\boldsymbol{x}) - A(\boldsymbol{\theta})\}, \tag{4.126}$$

is called an **exponential family**. The function ***T***(*x*) is called the **sufficient statistic**.

$$f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\eta}) = h(\boldsymbol{x})\exp\{\boldsymbol{\eta}^\top\boldsymbol{T}(\boldsymbol{x}) - A(\boldsymbol{\eta})\}, \tag{4.127}$$

is called the **canonical** (or **natural**) **exponential family**.

- The exponential family of distributions includes the exponential, gamma, normal, Poisson, binomial distributions, etc.

Example 4.3: Normal distribution. Consider a normal RV $X \sim N(\mu, \sigma^2)$. With $\boldsymbol{\theta} = (\mu, \sigma)$ we write the PDF of each sample $x_i \ \ (i = 1, 2, \ldots)$ as

$$
\begin{aligned}
f(x_i; \boldsymbol{\theta}) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_i^2}{2\sigma^2} + \frac{x_i \mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \log\sigma\right), \ \ i = 1, 2, \ldots, n.
\end{aligned}
$$

As in the previous example, we can present the normal distribution in the canonical exponential family form by identifying

$$
\boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} \\ \frac{\mu}{\sigma^2} \end{bmatrix}, \ \ \boldsymbol{T}(x) = \begin{bmatrix} -\frac{x^2}{2} \\ x \end{bmatrix},
$$

$$
h(\boldsymbol{X}) = \frac{1}{\sqrt{2\pi}}, \ \ A(\boldsymbol{\eta}) = \frac{\mu^2}{2\sigma^2} + \log\sigma.
$$

We can write the original parameter as $\boldsymbol{\theta} = (\mu, \sigma^2)$, where $\mu = \frac{\eta_2}{\eta_1}$ and $\sigma^2 = \frac{1}{\eta_1}$. Hence,

$$
A(\boldsymbol{\eta}) = \frac{\eta_2^2}{2\eta_1} - \frac{\log \eta_1}{2}.
$$

4.5 Bayesian Inference and Conjugate Priors

- Suppose that an observed sample *X* is drawn from a certain family of distributions specified by parameter θ.
- The Bayesian treats this parameter as a RV Θ, which is assigned a **prior** PDF $\boldsymbol{\pi(\theta)=f_{\Theta}(\theta)}$.
- If RV *X* is a discrete RV, we have from Bayes' theorem (2.63)

$$\boxed{\pi(\theta|x) = \frac{p(x|\theta)\pi(\theta)}{p(x)},} \tag{4.133}$$

where $p(x) = \sum_{\theta} p(x|\theta)\pi(\theta)$. $p(x) = \int_{\theta} p(x|\theta)\pi(\theta)d\theta$.

- If the RV *X* is a continuous RV,

$$\boxed{\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{f(x)},} \tag{4.134}$$

where $f(x) = \int_{\theta} f(x|\theta)\pi(\theta)\, d\theta$. $f(x) = \sum_{\theta} f(x|\theta)\pi(\theta)$

- The conditional PDF $f(x|\theta)$ is called the **likelihood function**, when it is viewed as a function of θ with given x, and is denoted as

$$L_x(\theta) = f(x|\theta) \text{ or } L_x(\theta) = p(x|\theta), \tag{4.135}$$

- Then the posterior distribution can be written as

$$\pi(\theta|x) \propto L_x(\theta)\pi(\theta). \tag{4.137}$$

- For certain choices of the prior distribution, the posterior distribution has the same mathematical form as the prior distribution. Such prior distribution is called a **conjugate prior** (**distribution**) of the given likelihood function.

Example 4.4: The Bernoulli distribution and its conjugate prior, the beta distribution

- Write the probability of success as θ (instead of *p*).
- Define the binary variable X_i which takes on 1 or 0, depending on the *i*th trial is a success (s) or failure (f).
- Then, we can write $p(x_i|\theta) = \theta^{x_i}(1-\theta)^{1-x_i}.$
- For n independent trials we observe the data $\boldsymbol{x} \triangleq (x_1, x_2, \ldots, x_n)^\top$

The likelihood function of θ given ***x*** is

$$L_{\boldsymbol{x}}(\theta) = p(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} p(x_i|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}$$
$$= \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}. \qquad (4.139)$$

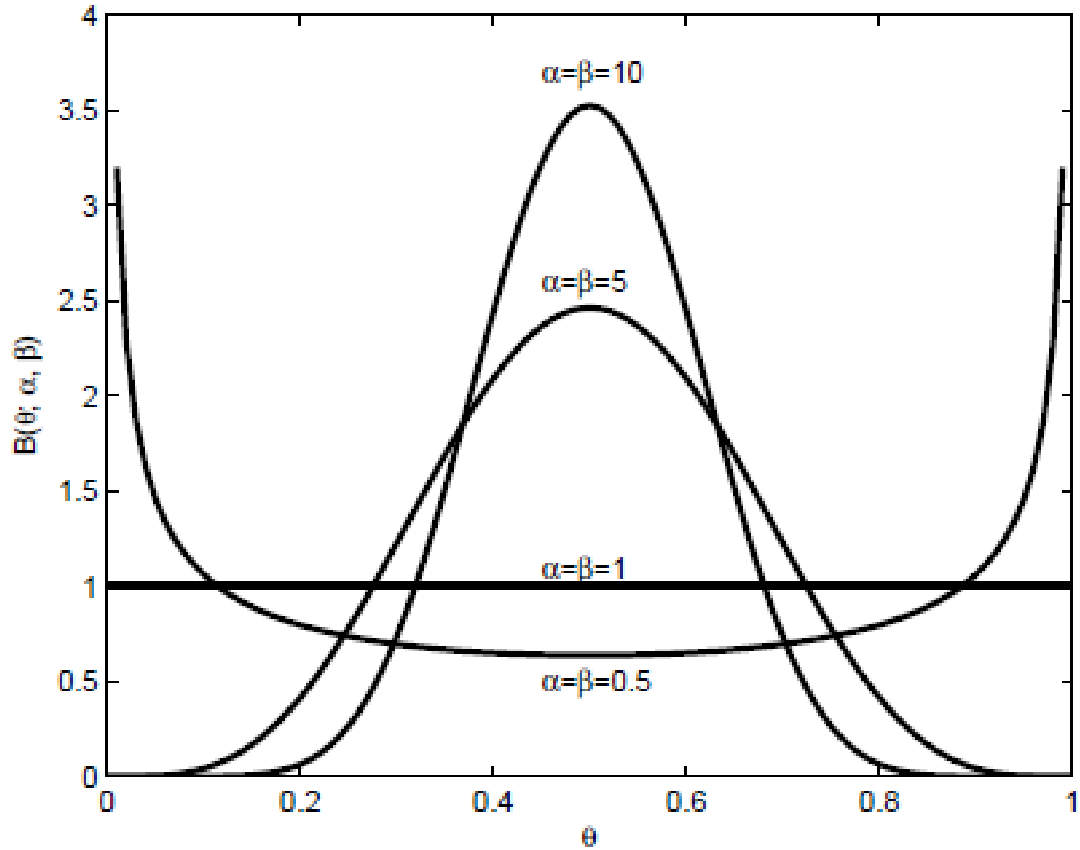
- As a prior distribution, consider the **beta distribution**:

$$\pi(\theta) = \mathrm{Beta}(\theta; \alpha, \beta) \triangleq \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}, \ \ 0 \le \theta \le 1, \ \alpha > 0, \ \beta > 0, \qquad (4.140)$$

where

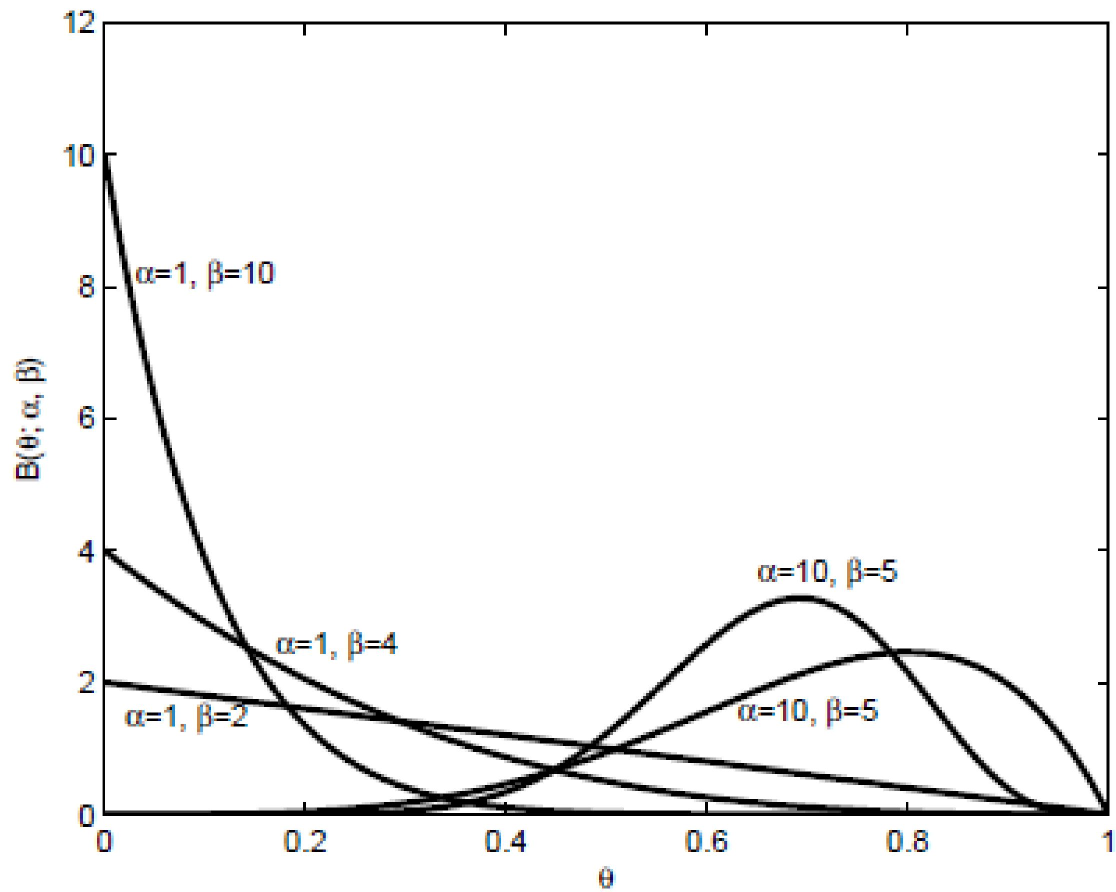
$$B(\alpha, \beta) = \int_0^\infty \theta^{\alpha-1}(1-\theta)^{\beta-1}\, d\theta \qquad (4.141)$$

α and β are called **prior hyperparameters** (cf, the model parameter θ).



(a)

Figure 4.8 The PDF of beta distribution $\mathrm{Beta}(\theta; \alpha, \beta)$ of (4.140) for (a) $\alpha = \beta = 0.5, 1.0, 5$ and 10;



(b)

Figure 4.8 The PDF of beta distribution $\text{Beta}(\theta; \alpha, \beta)$ of (4.140) for

(b) $(\alpha, \beta) = (1, 2), (1, 4), (1, 10), (10, 5)$ and $(5, 2)$.

Note: The rightmost curve corresponds to (5, 2)

❖ The **beta function** is related to the **gamma function** (see (4.31) of p. 78)

$$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \tag{4.142}$$

❖ The mean and variance of this prior distribution are

$$E[\Theta]=\frac{\alpha}{\alpha+\beta}, \text{ and } \mathrm{Var}[\Theta]=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{4.143}$$

❖ The posterior probability can be evaluated as

$$\begin{aligned}\pi(\theta|\boldsymbol{x}) &\propto p(\boldsymbol{x}|\theta)\pi(\theta) \propto \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}\theta^{\alpha-1}(1-\theta)^{\beta-1}\\ &\propto \theta^{(\alpha+\sum_{i=1}^n x_i)-1}(1-\theta)^{(\beta+n-\sum_{i=1}^n x_i)-1},\end{aligned} \tag{4.144}$$

❖ Thus, the posterior probability is also a beta distribution Beta(θ; α_1, β_1),

$$\alpha_1=\alpha+\sum_{i=1}^n x_i, \text{ and } \beta_1=\beta+n-\sum_{i=1}^n x_i, \tag{4.145}$$

$$\begin{aligned}E[\Theta|\boldsymbol{x}] &= \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)\frac{\alpha}{\alpha+\beta}+\left(\frac{n}{\alpha+\beta+n}\right)\overline{x}_n,\\ &= \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)E[\Theta]+\left(\frac{n}{\alpha+\beta+n}\right)\hat{\theta}_{\mathrm{MLE}}(\boldsymbol{x}),\end{aligned} \tag{4.146}$$

where we call α_1 and β_1 the **posterior hyperparameters**, and

$$\hat{\theta}_{\mathrm{MLE}}(x) = \overline{x}_n \triangleq \frac{x_1 + x_2 + \ldots + x_n}{n}$$

is the **maximum likelihood estimate (MLE)** of θ, which is the value that maximizes the likelihood function $L_x(\theta)$ of (4.139).

- ❖ As the sample size n increases, the weight on the prior means diminishes, whereas the weight on the MLE approaches one. This behavior illustrates how **Bayesian inference** generally works.
- ❖ For a likelihood function that belongs to the **exponential family**, i.e.,

$$L_{\boldsymbol{x}}(\boldsymbol{\theta}) = h(\boldsymbol{x}) \exp\{\boldsymbol{\eta}^\top(\boldsymbol{\theta})\boldsymbol{T}(\boldsymbol{x}) - A(\boldsymbol{\theta})\}, \quad (4.147)$$

conjugate priors can be constructed as follows:

$$f(\boldsymbol{\theta}; \boldsymbol{\alpha}, \beta) \propto \exp\{\boldsymbol{\eta}^\top(\boldsymbol{\theta})\boldsymbol{\alpha} - \beta A(\boldsymbol{\theta})\}, \quad (4.148)$$

then the posterior distribution takes the form

$$f(\boldsymbol{\theta}|\boldsymbol{x}; \boldsymbol{\alpha}, \beta) \propto \exp\{\boldsymbol{\eta}^\top(\boldsymbol{\theta})[\boldsymbol{\alpha} + \boldsymbol{T}(\boldsymbol{x})] - (1+\beta)A(\boldsymbol{\theta})\}, \quad (4.149)$$

i.e., $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha} + \boldsymbol{T}(\mathbf{x})$, and $\beta_1 = 1 + \beta$.

5 Functions of Random Variables and Their Distributions

5.1 Functions of One Random Variable

Consider Y=g(X), where X is a RV and $g(\cdot)$ is a mapping from **R** to **R**.

Then Y is also a RV with

$$F_Y(y) = P[Y \leq y] = P[g(X) \leq y] \\ = P[X \in \mathcal{D}_y], \quad (5.2)$$

where

$$\mathcal{D}_y = \{x : \; g(x) \leq y\}. \quad (5.3)$$

Then

$$\boxed{F_Y(y) = \int_{-\infty}^{\infty} I(x \in \mathcal{D}_y) f_X(x)\, dx} \quad (5.4)$$

where

$$I(A) = \begin{cases} 1, & \text{if } A \text{ is true;} \\ 0, & \text{otherwise.} \end{cases} \quad (5.5)$$

Example 4.2 Square law detector.

Consider $Y=g(X)=X^2$. Then $\mathcal{D}_y = [-\sqrt{y}, \sqrt{y}\,]$

$$F_Y(y) = P[X^2 \leq y] = P[-\sqrt{y} \leq X \leq \sqrt{y}] = F_X(\sqrt{y}) - F_X(-\sqrt{y}) \text{ for } y \geq 0. \qquad (5.11)$$

By differentiating this,

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})]. \qquad (5.12)$$

An alternative way to derive the above PDF: Note that $y=x^2$ has two solutions $x_1= \sqrt{y}$ and $x_2= -\sqrt{y}$

Then,

$$P[y < Y \leq y + \delta y] = P[x_1 < X \leq x_1 + \delta x_1] + P[x_2 + \delta x_2 \leq X < x_2]. \qquad (5.13)$$

$$f_Y(y)\delta y \approx f_X(x_1)\delta x_1 + f_X(x_2)(-\delta x_2). \qquad (5.14)$$

$$f_Y(y) = \frac{f_X(x_1)}{g'(x_1)} + \frac{f_X(x_2)}{-g'(x_2)}. \qquad (5.15)$$

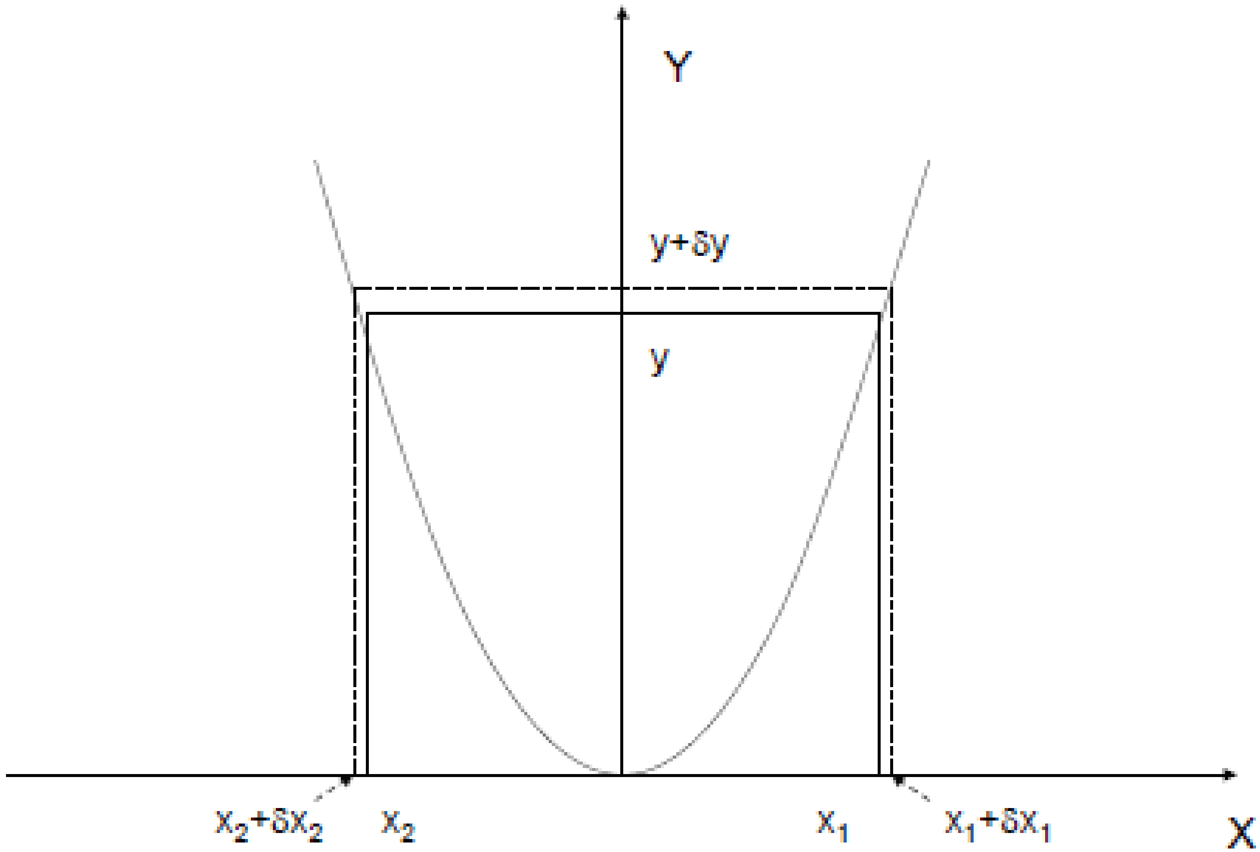


Figure 5.1 The nonlinear function $Y = X^2$ and mapping of the interval $(y, y+\delta y)$ into the two disjoint intervals $(x_1, x_1+\delta x_1)$ and $(x_2+\delta x_2, x_2)$ $(\delta x_2 < 0)$.

❖ **Generalization of the previous example**:

Suppose that for given *y*, *y*=*g*(*x*) has multiple solutions $x_1, x_2, \ldots, x_m$, where the number of solutions, *m*, depends on *y*. So we write it as *m*(*y*).

$$x_i = g^{-1}(y), \quad i = 1, 2, \ldots, m(y), \tag{5.16}$$

If *g*(*x*) is continuous at all these *m*(*y*) points, then

$$f_Y(y) = \sum_{i=1}^{m(y)} \frac{f_X(x_i)}{|\, g'(x_i)\, |}. \tag{5.17}$$

5.2 Function of Two Random Variables

Let

$$Z = g(X,Y). \tag{5.18}$$

Then

$$\begin{aligned} F_Z(z) &= P[g(X,Y) \leq z] = P[(X,Y) \in \mathcal{D}_z] \\ &= \int\int I((x,y) \in \mathcal{D}_z) f_{XY}(x,y)\,dx\,dy, \end{aligned} \tag{5.19}$$

where

$$\mathcal{D}_z = \{(x,y) : \ g(x,y) \leq z\}, \tag{5.20}$$

Example 5.3: Sum of two RVs:

Consider Z=X+Y. Then $\mathcal{D}_z = \{(X,Y) : X+Y \leq z\}$

We can represent $\mathcal{D}_z = \bigcup_{-\infty<y<\infty} \mathcal{H}_y$.

where

$\mathcal{H}_y$ ={(*X*, *Y*): *y*<*Y*<*y*+*dy*, - ∞<*X*<*z*-*y*}

is a horizontal strip of width *dy*.

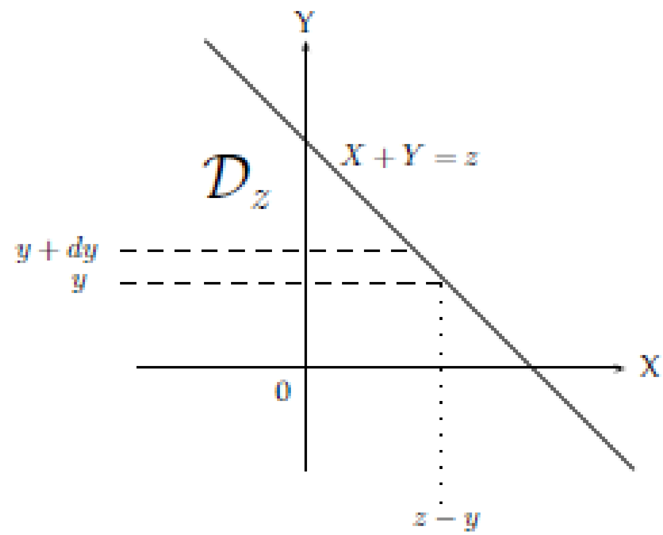


Figure 5.2 The region $\mathcal{D}_z = \{(X,Y) : X+Y \le z\}$ and a horizontal strip of width dy extending horizontally over the interval $\{-\infty < X < z-y\}$.

Thus,

$$F_Z(z) = \int\int I(x+y \le z) f_{XY}(x,y)\,dxdy = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{z-y} f_{XY}(x,y)\,dx\right] dy \qquad (5.21)$$

$$f_Z(z) = \int_{-\infty}^{\infty}\left[\frac{\partial}{\partial z}\int_{-\infty}^{z-y} f_{XY}(x,y)\,dx\right] dy. \qquad (5.22)$$

Consider Leibniz's rule (5.94):

$$\boxed{\frac{d}{dz}\int_{a(z)}^{b(z)} h(z,y)\,dy = h(z,b(z))b'(z) - h(z,a(z))a'(z) + \int_{a(z)}^{b(z)} \frac{\partial}{\partial z} h(z,y)\,dy.} \tag{5.94}$$

Then

$$\begin{aligned}\frac{\partial}{\partial z}\int_{-\infty}^{z-y} f_{XY}(x,y)\,dx &= f_{XY}(z-y,y)\cdot 1 - f_{XY}(-\infty,y)\cdot 0 + \int_{-\infty}^{z-y} \frac{\partial f_{XY}(x,y)}{\partial z}\,dx \\ &= f_{XY}(z-y,y).\end{aligned} \tag{5.23}$$

Thus,

$$\boxed{f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(z-y,y)\,dy.} \tag{5.24}$$

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x,z-x)\,dx. \tag{5.25}$$

- If X and Y are independent,

$$f_{XY}(x,y) = f_X(x) f_Y(y), \tag{5.26}$$

then

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx. \tag{5.27}$$

$$\boxed{f_Z(z) = f_X(z) \circledast f_Y(z).} \tag{5.28}$$

$$\boxed{F_Z(z) = \int_{-\infty}^{\infty} F_X(z-y)\, dF_Y(y) = \int_{-\infty}^{\infty} F_Y(z-x)\, dF_X(x),} \tag{5.29}$$

$$\boxed{F_Z(z) = F_X(z) \circledast f_Y(z) = f_X(z) \circledast F_Y(z).} \tag{5.30}$$

5.3 Two Functions of Two Random Variables and the Jacobian Matrix

$$U = g(X,Y), \text{ and } V = h(X,Y). \tag{5.38}$$

$$\mathcal{D}_{u,v} = \{(x,y):\ g(x,y) \le u, h(x,y) \le v\}. \tag{5.39}$$

$$\begin{aligned} F_{UV}(u,v) &= P[g(X,Y) \le u, h(X,Y) \le v] = P[(X,Y) \in \mathcal{D}_{u,v}] \\ &= \int\int I((x,y) \in \mathcal{D}_{u,v}) f_{XY}(x,y)\,dx\,dy, \end{aligned} \tag{5.40}$$

Assume that $g(x, y)$ and $h(x, y)$ are continuous and differentiable functions. Given $(U, V)=(u,v)$, there are multiple solutions $(X,Y)=(x_i, y_i)$, i=1, 2, …, m such that

$$u = g(x_i, y_i), \text{ and } v = h(x_i, y_i),\ i = 1, 2, \ldots, m. \tag{5.41}$$

Let the inverse mapping be

$$x_i = p_i(u,v), \text{ and } y_i = q_i(u,v). \tag{5.42}$$

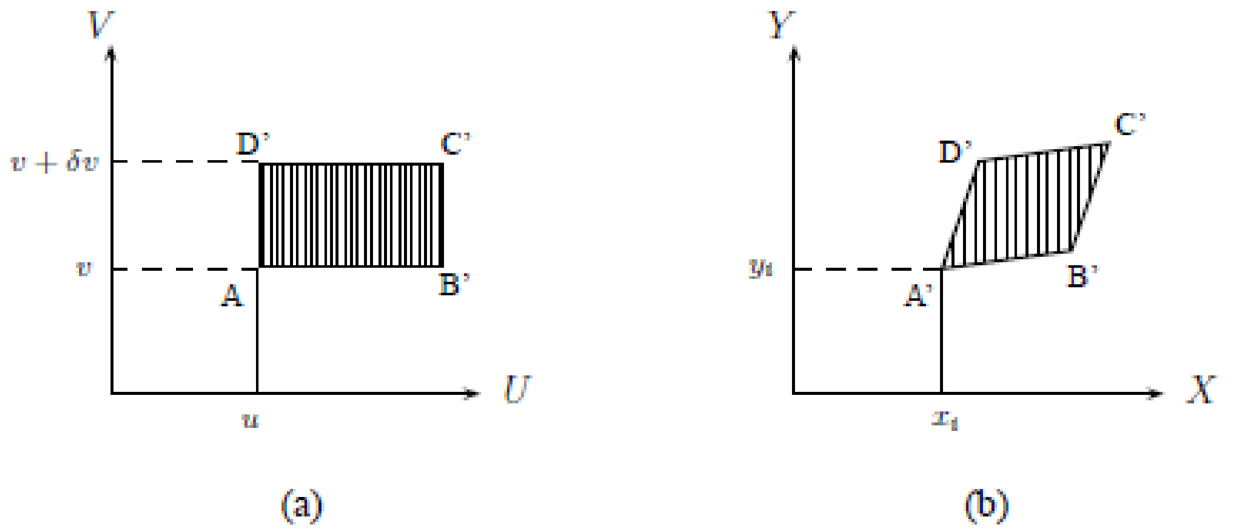


Figure 5.4 (a) The rectangle $ABCD$ at (u, v) in the U-V plane; (b) the i-th parallelogram $A'B'C'D'$ at (x_i, y_i) in the X-Y plane, where $x_i = p_i(u, v)$ and $y_i = q_i(u, v)$.

$$\begin{aligned}
A &= (u, v) & &\to A' = (x_i, y_i);\\
B &= (u + \delta u, v) & &\to B' = \left(x_i + \frac{\partial}{\partial u} p_i(u, v)\delta u, y_i + \frac{\partial}{\partial u} q_i(u, v)\delta u\right);\\
C &= (u, v + \delta v) & &\to C' = \left(x_i + \frac{\partial}{\partial v} p_i(u, v)\delta v, y_i + \frac{\partial}{\partial v} q_i(u, v)\delta v\right);\\
D &= (u + \delta u, v + \delta v) & &\to D' = C' + (B' - A').
\end{aligned} \tag{5.43}$$

Note: In the above figure (a) B', C' and D' should be labeled as B, D and C, respectively. In (b), C' and D' should be labeled as D' and C', respectively.

The probability that (*U*, *V*) falls in the rectangular *ABCD*:

$$P[u < U \leq u + \delta u, v < V \leq v + \delta v] = f_{UV}(u, v)\delta u\, \delta v, \tag{5.44}$$

$$= \sum_{i=1}^{m} f_{XY}(x_i, y_i)\Delta_i, \tag{5.45}$$

where Δ_i is the area A'B'C'D'.

Recall the formula (Problem 5.17) for the area S of a triangular defined by (x_1, y_1) , (x_2, y_2) and (x_3, y_3)

$$S = \left| \frac{1}{2} \det \begin{bmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix} \right|. \tag{5.46}$$

Then

$$\frac{\Delta_i}{2} = \left| \frac{1}{2} \det \begin{bmatrix} \frac{\partial}{\partial u} p_i(u, v)\delta u & \frac{\partial}{\partial v} p_i(u, v)\delta v \\ \frac{\partial}{\partial u} q_i(u, v)\delta u & \frac{\partial}{\partial v} q_i(u, v)\delta v \end{bmatrix} \right|$$

$$= \left| \frac{1}{2} \det \begin{bmatrix} \frac{\partial}{\partial u} p_i(u, v) & \frac{\partial}{\partial v} p_i(u, v) \\ \frac{\partial}{\partial u} q_i(u, v) & \frac{\partial}{\partial v} q_i(u, v) \end{bmatrix} \delta u\, \delta v \right|. \tag{5.48}$$

Define the Jacobian matrix of the mapping $p_i(\underline{u}, v)$ and $q_i(u, v)$:

$$\boldsymbol{J}\left(\frac{p_i, q_i}{u, v}\right) \triangleq \begin{bmatrix} \frac{\partial p_i}{\partial u} & \frac{\partial p_i}{\partial v} \\ \frac{\partial q_i}{\partial u} & \frac{\partial q_i}{\partial v} \end{bmatrix}, \tag{5.49}$$

Then

$$f_{UV}(u, v) = \sum_{i=1}^{m} \left| \boldsymbol{J}\left(\frac{p_i, q_i}{u, v}\right) \right| f_{XY}(x_i, y_i), \tag{5.52}$$

The determinant det $\boldsymbol{J}$ is called the Jacobian or Jacobian determinant. If we define the Jacobian matrix of the original mapping by

$$\boldsymbol{J}\left(\frac{g, h}{x, y}\right) = \begin{bmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{bmatrix}, \tag{5.53}$$

then

$$\left| \boldsymbol{J}\left(\frac{p_i, q_i}{u, v}\right) \right| = \left| \boldsymbol{J}\left(\frac{g, h}{x, y}\right) \right|^{-1}. \tag{5.54}$$

Example 5.6: Two linear transformations.

$g(X, Y)=aX+bY$ and $h(X, Y)= cX + dY$, $(ad\text{-}bc\neq 0)$

$$\begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}. \quad (5.56)$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}. \quad (5.57)$$

Thus,

$$p_1(u,v) = \frac{1}{\Delta}(du - bv), \text{ and } q_1(u,v) = \frac{1}{\Delta}(-cu + av), \quad (5.58)$$

where

$$\Delta = ad - bc. \quad (5.59)$$

$$f_{UV}(u,v) = \left| J\left(\frac{p_1, q_1}{u, v}\right)\right| f_{XY}(x_1, y_1)$$
$$= |\Delta|^{-1} f_{XY}\left(\Delta^{-1}(du - bv), \Delta^{-1}(-cu + av)\right). \quad (5.62)$$

❖ Consider a special case, *a*=*b*=*c*=1 and *d*=0, i.e., *U*=*X* + *Y* and *V*=*X*. Then

$$f_{UV}(u,v) = |-1| f_{XY}(v, u-v). \tag{5.64}$$

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(v, u-v)\, dv, \tag{5.65}$$

❖ If we set *a*=*b*=*d*=1 and *c*=0, i.e., *U*=*X*+*Y* and *V*=*Y*. Then

$$f_{UV}(u,v) = f_{XY}(u-v, v), \tag{5.66}$$

$$f_U(u) = \int_{-\infty}^{\infty} f_{XY}(u-v, v)\, dv, \tag{5.67}$$