4. Families of Distributions and Transformations

1. Common Families of Distributions

In this section, we introduce some commonly used families of distributions. The index (or indices) used to denote a member of a family is called the *parameter*.

Uniform distribution, denoted by Uniform $[a, b]$, is given by a family of densities

$$p_{a,b}(x) = \frac{1}{b-a}\,\mathrm{I}\{a \le x \le b\}$$

with respect to the Lebesque measure, where the parameters a and b are the end points of an interval. The random experiment of selecting a number randomly over the interval $[a, b]$, and the random variable assigning the selected number induces such distribution.

Bernoulli distribution, denoted by Bernoulli (θ) for $\theta \in [0, 1]$, is given by a family of densities

$$p_\theta(x) = \theta^x(1-\theta)^{1-x}\,\mathrm{I}\{x = 0, 1\}$$

with respect to the counting measure. The distribution is induced by a random variable taking values 1 and 0, respectively for the outcomes perceived as 'success' and 'failure', for a random experiment such as toss of a coin. The parameter θ can then be interpreted as the probability of success.

Binomial distribution, Binomial (n, θ) for a positive integer n and $\theta \in [0, 1]$, is given by a family of densities

$$p_{n,\theta}(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}\,\mathrm{I}\{x = 0, 1, \ldots, n\}$$

The typical random variable inducing such distribution is that assigns the number of heads in an n consecutive tosses of a coin. In general, we may interpret the parameter n as the number of trials, and the parameter θ as the probability of success for each trial.

Remark We can represent a binomial random variable X as a sum of independent and identically distributed (i.i.d) Bernoulli random variables. For X, which assigns the number of success to each outcome in an n-consecutive trials, we write

$$X = Y_1 + \cdots + Y_n$$

where Y_i, $i = 1, \ldots, n$, is a random variable taking value 1 if the i-th trial is success, and 0 if it is failure. It is clear that Y_i's are independent, and has the common distribution Bernoulli (θ). Mean, variance and mgf of Binomial (n, θ) can thus be easily obtained from those of Bernoulli (θ).

Poisson distribution, Poisson (λ) with $\lambda > 0$, has the density

$$p_\lambda(x) = e^{-\lambda}\frac{\lambda^x}{x!}\,\mathrm{I}\{x = 0, 1, 2, \ldots\}$$

The parameter λ represents both mean and variance.

Remarks

(a) The typical Poisson random variable arises in the context of observing the occurrence of certain happenings in time, space, region or length, such as the arrival of a telephone call, a traffic accident, etc. To show that such a happening can be appropriately modelled by a Poisson random variable, let X denote the number of telephone calls in a time interval. Suppose that the expected number of calls in the interval is λ. Now divide the interval into n small sub-intervals, with n large enough that no two telephone calls arrive in a sub-interval. Then we may regard X as a binomial random variable with parameters $(n, \lambda/n)$. The density of X at x is then given by

$$\begin{aligned}\binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} &= \frac{(n)_x}{n^x}\left(1-\frac{\lambda}{n}\right)^{-x}\left(1-\frac{\lambda}{n}\right)^{n}\frac{\lambda^x}{x!} \\ &\to e^{-\lambda}\frac{\lambda^x}{x!}\end{aligned}$$

as $n \to \infty$, and where we used the notation $(n)_x = n!/(n-x)!$ and the facts

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}, \quad \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-x} = 1 \quad \text{and} \quad \lim_{n\to\infty}\frac{(n)_x}{n^x} = 1$$

(b) In dealing with Poisson density, the power series expansion

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$$

of the exponential function is used repeatedly.

Normal distribution, $\mathbf{N}(\mu, \sigma^2)$, or Gaussian distribution has the density

$$p_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

It is the most important distribution in probability and statistics. The parameters μ and $\sigma^2 > 0$ are, respectively, the mean and variance. The normal distribution with mean 0 and variance 1, $\mathbf{N}(0,1)$, is called the *standard normal* distribution. It was originally discovered as well fitted for the distribution of observation errors.

Gamma distribution, Gamma (r, λ), is represented by a family of densities

$$p_{r,\lambda}(x) = \frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1} e^{-\lambda x}\, \mathrm{I}\{x \geq 0\}$$

where $r, \lambda > 0$ are the parameters, and $\Gamma(\cdot)$ is the *gamma function* defined by

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt$$

Lemma 1

$$\Gamma(z+1) = z\Gamma(z) \quad \text{and} \quad \Gamma(1) = 1$$

Proof Notice by integration by parts that

$$\begin{aligned} \Gamma(z+1) &= \int_0^\infty t^z e^{-t} dt \\ &= -t^z e^{-t}\Big|_0^\infty + z\int_0^\infty t^{z-1} e^{-t} dt \\ &= z\Gamma(z) \end{aligned}$$

Moreover,

$$\Gamma(1) = \int_0^\infty e^{-t} dt = 1$$

as required. ■

Remark Note that $\Gamma(n+1) = n!$ for a positive integer n. The gamma function thus generalizes the factorial function.

Remark The family of gamma distributions has two important sub-family, *exponential distribution*, Exponential (λ) = Gamma $(1, \lambda)$, and *chi-square distribution*, $\chi^2(k)$ = Gamma $(k/2, 1/2)$.

Beta distribution, Beta(a, b), is given by the density

$$p_{a,b}(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}\,\mathrm{I}\{0 \le x \le 1\}$$

where $B(a, b)$ denotes the beta function defined by

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx$$

Remarks

(a) The beta function is related to the gamma function through the identity

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

(b) The beta distribution reduces to Uniform$[0, 1]$ when $a = b = 1$.

Cauchy distribution, Cauchy (a, b), has the density given by

$$p_{a,b}(x) = \frac{1}{\pi b \left(1 + \left(\frac{x-a}{b}\right)^2\right)}$$

where $-\infty < a < \infty$ and $b > 0$ are parameters. The shape of Cauchy density is quite similar to that of normal density. However, it has no finite moments.

Multinomial distribution Let $X = (X_1, \ldots, X_m)'$. Then *multinomial distribution* is defined by

$$p(x_1, \ldots, x_m; p_1, \ldots, p_m) = \frac{n!}{x_1! \cdots x_m!} p_1^{x_1} \cdots p_m^{x_m}\,\mathrm{I}\left\{\sum_{k=1}^{m} x_k = n;\; x_k \ge 0\right\}$$

The typical setting of multinomial distribution is n consecutive rolls of a die. In this case $m = 6$ and $X_j(j = 1, \ldots, 6)$ denotes the number of times j spots up in n consecutive rolls of a die. Note that we can represent the multinomial distribution by binomial distribution. We will illustrate this using the example of n consecutive rolls of a die. Let $X = \sum_{i=1}^{n} Y_i$ where Y_i is 6-dimensional random vector. If, for example, 3 spots were up in the i-th roll, then $Y_i(\omega)$ takes the form

$$Y_i(\omega) = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Notice that, for each $k = 1, \ldots, 6$, $\{Y_{ki}\}_{i=1}^{n}$ are i.i.d. Bernoulli random variable with parameter p_k, which implies $X_k = \sum_{i=1}^{n} Y_{ki}$ is Binomial(n, p_k) for $k = 1, \ldots, 6$.

Mean, Variance and Moment Generating Function

distribution	mean	variance	mgf
Uniform $[a, b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt} - e^{at}}{(b-a)t}$
Bernoulli (θ)	θ	$\theta(1-\theta)$	$(1-\theta) + \theta\, e^t$
Poisson (λ)	λ	λ	$\exp\Big(\lambda(e^t - 1)\Big)$
Normal (μ, σ^2)	μ	σ^2	$\exp\left(\mu\iota + \frac{\sigma^2 t^2}{2}\right)$
Gamma (r, λ)	$\frac{r}{\lambda}$	$\frac{r}{\lambda^2}$	$\left(\frac{\lambda}{\lambda - t}\right)^r$
Beta (a, b)	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	$\cdot$

2. Distributions of Functions of Random Variables

Let the distribution of a random variable or a random vector X be known. We show in this section how to obtain the distribution of a random variable Y defined by $Y = f(X)$ for a measurable f. There are three commonly used methods — *distribution function technique, moment generating function technique*, and *transformation technique.*

Distribution Function Technique It is sometimes possible to compute the distribution function G of Y directly from

$$G(y) = \mathbf{P}\{Y \leq y\} = \mathbf{P}\{f(X) \leq y\}$$

as we see in the following examples.

Examples:

(a) Let $X \sim U[0,1]$ and $Y = -\log X$. Then the distribution function G of Y is

$$\begin{aligned} G(y) &= \mathbf{P}\{-\log X \leq y\} \\ &= \mathbf{P}\{X \geq e^{-y}\} \\ &= 1 - e^{-y} \end{aligned}$$

when $y \geq 0$, and 0 otherwise. Note that the distribution function F of X is given by $F(x) = x$ and 1, respectively for $0 \leq x < 1$ and $x \geq 1$.

(b) Let X_i be independent random variables with distribution functions F_i, $i = 1, \ldots, n$. Define $Y = \max\{X_1, \ldots, X_n\}$. Then the distribution function G of Y is given by

$$\begin{aligned} G(y) &= \mathbf{P}\left(\bigcap_{i=1}^{n} \{X_i \leq y\}\right) \\ &= \prod_{i=1}^{n} \mathbf{P}\{X_i \leq y\} \\ &= \prod_{i=1}^{n} F_i(y) \end{aligned}$$

If X_i, $i = 1, \ldots, n$, are i.i.d. with the common distribution function F, then $G(y) = F(y)^n$.

(c) Let $X = (X_1, X_2)'$ be a random vector with distribution P and density p. Then the distribution function G of $Y = X_1 + X_2$ is given by

$$\begin{aligned} G(y) &= \mathbf{P}\{X_1 + X_2 \leq y\} \\ &= \mathrm{P}\{(x_1, x_2) |\, x_1 + x_2 \leq y\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{y - x_2} p(x_1, x_2)\, dx_1 dx_2 \end{aligned}$$

which can easily be obtained if p is specified.

Moment Generating Function Technique The distribution of sums of *independent* random variables can readily be obtained via moment generating functions, as they uniquely define distributions. If we let $X_1, \ldots, X_n$ be independent random variables with moment generating functions m_i, $i = 1, \ldots, n$, and $Y = X_1 + \cdots + X_n$, then the moment generating function m of Y is given by

$$m(t) = \mathbf{E}(e^{t(X_1 + \cdots + X_n)}) = \prod_{i=1}^{n} m_i(t)$$

Examples:

(a) Let $X_i \sim \text{Poisson}(\lambda_i)$, $i = 1, \ldots, n$, be *independent*. Then the moment generating function of $Y = X_1 + \cdots + X_n$ is

$$m(t) = \prod_{i=1}^{n} \exp\Big(\lambda_i (e^t - 1)\Big) = \exp\left((e^t - 1) \sum_{i=1}^{n} \lambda_i\right)$$

from which we may deduce that Y has Poisson distribution with parameter $\sum \lambda_i$.

(b) Let $X_i \sim \mathbf{N}(\mu_i, \sigma_i^2)$ be independent. The moment generating function of $Y = c_1 X_1 + \cdots + c_n X_n$ is given by

$$\begin{aligned} m(t) &= \prod_{i=1}^{n} \exp\left(c_i \mu_i t + \frac{c_i^2 \sigma_i^2 t^2}{2}\right) \\ &= \exp\left(t \sum_{i=1}^{n} c_i \mu_i + \frac{t^2}{2} \sum_{i=1}^{n} c_i^2 \sigma_i^2\right) \end{aligned}$$

which implies that

$$Y \sim \mathbf{N}\left(\sum_{i=1}^{n} c_i\mu_i, \sum_{i=1}^{n} c_i^2\sigma_i^2\right)$$

Transformation Technique If the function f is one-to-one, we may find the density of $Y = f(X)$ from that of X by the transformation technique. To introduce the method, we first let f be a measurable, one-to-one transformation in $\mathbf{R}^n$. Write $f = (f_1, \ldots, f_n)'$ and $x = (x_1, \ldots, x_n)'$. Also, denote by P_X and P_Y the distributions of X and Y, respectively, having densities p_X and p_Y with respect to the counting or the Lebesque measure μ on $\mathbf{R}^n$.

We now let $B \in \mathcal{B}(\mathbf{R}^n)$ be chosen arbitrarily. Then define $A = f^{-1}(B)$. Note that $A \in \mathcal{B}(\mathbf{R}^n)$ since f is assumed to be measurable, and that $\{X \in A\} = \{Y \in B\}$. We therefore have

$$P_Y(B) = P_X(A) = \int_A p_X(x)\, d\mu(x)$$

When μ is the counting measure, we immediately have

$$\int_A p_X(x)\, d\mu(x) = \sum_{x \in A} p_X(x) = \sum_{y \in B} p_X(f^{-1}(y))$$

If μ is the Lebesque measure and f is differentiable, we have from the change of variables formula for the transformation $y = f(x)$ that

$$\int_A p_X(x)\, d\mu(x) = \int_A p_X(x)\, dx = \int_B p_X(f^{-1}(y)) \left| \det \dot{f}(f^{-1}(y)) \right|^{-1} dy$$

where $\dot{f}$ is the Jacobian matrix of f, i.e., the matrix $(\partial f_i / \partial x_j)$ of the first derivatives of f. We may now easily deduce that the density p_Y of Y is given by

$$p_Y(y) = p_X(f^{-1}(y)) \;\text{ or }\; p_X(f^{-1}(y)) \left| \det \dot{f}(f^{-1}(y)) \right|^{-1}$$

depending upon whether X and Y have densities with respect to the counting or Lebesque measure.

Obtaining p_Y from p_X is thus extremely simple when they are densities with respect to the counting measure. We just replace x by $f^{-1}(y)$ in the argument of p_X.

If, however, they are the densities with respect to the Lebesque measure, we need to make an additional adjustment for the change in the volume element from dx to dy. This can easily be done using the total differentiation formula

$$dy = \left|\det \dot{f}(x)\right| dx = \left|\det \dot{f}(f^{-1}(y))\right| dx$$

The adjustment factor

$$\left|\det \dot{f}(f^{-1}(y))\right|^{-1}$$

for the change in the volume elements is often called the *Jacobian of transformation.*

Example The joint desity of X_1 and X_2 is given by

$$\begin{aligned} p(x_1, x_2) &= 4x_1x_2 \;\; \text{if } 0 < x_1, x_2 < 1, \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

Define $Y_1 = X_1/X_2$ and $Y_2 = X_1X_2$. Now we will obtain the joint density of Y_1 and Y_2. Let $\mathcal{X} = \{(x_1, x_2)|\, 0 < x_1, x_2 < 1\}$, i.e., the support of the joint density of X_1 and X_2. Since the inverse transformation is

$$x_1 = \sqrt{y_1y_2} \quad \text{and} \quad x_2 = \sqrt{y_2/y_1}$$

the support $\mathcal{Y}$ of the joint density of Y_1 and Y_2 is given by $\mathcal{Y} = \{(y_1, y_2)|y_1, y_2 > 0, y_1y_2 < 1, y_2 < y_1\}$.

We now concentrate on the transformation on $\mathcal{X}$ and $\mathcal{Y}$. We have from the total differentiation

$$\begin{pmatrix} dx_1 \\ dx_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\sqrt{y_2/y_1} & \frac{1}{2}\sqrt{y_1/y_2} \\ -\frac{1}{2}\sqrt{y_2/y_1^3} & \frac{1}{2}\sqrt{1/y_1y_2} \end{pmatrix} \begin{pmatrix} dy_1 \\ dy_2 \end{pmatrix}$$

which yields by the change of variables formula

$$dx_1\, dx_2 = \left|\det \begin{pmatrix} \frac{1}{2}\sqrt{y_2/y_1} & \frac{1}{2}\sqrt{y_1/y_2} \\ -\frac{1}{2}\sqrt{y_2/y_1^3} & \frac{1}{2}\sqrt{1/y_1y_2} \end{pmatrix}\right| dy_1\, dy_2 = \frac{1}{2y_1}\, dy_1\, dy_2$$

Note that $y_1 > 0$ on $\mathcal{Y}$. Therefore, we have

$$p(x_1, x_2)\, dx_1\, dx_2 = 4x_1x_2\, dx_1\, dx_2 = \frac{4y_2}{2y_1}\, dy_1\, dy_2$$

The joint density of Y_1 and Y_2 is therefore by

$$p(y_1, y_2) = \frac{2y_2}{y_1}$$

over the support $\mathcal{Y}$.

3. Exercises

1. Find mean, variance and mgf of

$$\text{Uniform}[a, b],\ \text{Poisson}(\lambda),\ \text{Normal}(\mu, \sigma^2),\ \text{Gamma}(r, \lambda)$$

2. Let X and Y be independent standard normal variates. Find the density of a random variable defined by

$$U = \frac{X}{Y}$$

3. Let X and Y be independent random variables having, respectively, standard normal and chi-square distributions with k-degrees of freedom. Obtain the density of a random variable defined by

$$U = \frac{X}{\sqrt{Y/k}}$$

4. Let X and Y be independent random variables having, respectively, chi-square distributions with p- and q-degrees of freedom. Obtain the density of a random variable defined by

$$U = \frac{X/p}{Y/q}$$

Remark: For Problems 2 – 4, let $V = Y$ and first find the joint density of U and V.