Chapter 3

Common Families of Distributions

3.1 Discrete Distributions

A random variable X is said to have a discrete distribution if the range of X, the sample space, is countable. In most situations, the random variable has integer-valued outcomes.

3.1.1 Discrete Uniform Distribution

A random variable X has a discrete uniform $(1, N)$ distribution if

$$P(X = x|N) = \frac{1}{N}, \quad x = 1, 2, \ldots, N,$$

where N is a specified integer. This distribution puts equal mass on each of the outcomes $1, 2, \ldots, N$.

We then have

$$EX = \sum_{x=1}^{N} xP(X = x|N) = \sum_{x=1}^{N} x\frac{1}{N} = \frac{N+1}{2},$$

and

$$EX^2 = \sum_{x=1}^{N} x^2 \frac{1}{N} = \frac{(N+1)(2N+1)}{6},$$

and so,

$$\text{Var}X = EX^2 - (EX)^2 = \frac{(N+1)(2N+1)}{6} - (\frac{N+1}{2})^2 = \frac{(N+1)(N-1)}{12}.$$

3.1.2 Hypergeometric Distribution

Suppose we have a large urn filled with N balls that are identical in every way except that M are red and $N-M$ are green. We reach in, blindfolded, and select K balls at random. Let X denote the number of red balls in a sample of size K, then X has a *hypergeometric distribution* given by

$$P(X = x|N, M, K) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}, \quad x = 0, 1, \ldots, K.$$

The requirements for the range of X is

$$M \geq x \quad \text{and} \quad N - M \geq K - x,$$

which can be combined as

$$M - (N - K) \leq x \leq M.$$

The mean of this distribution is

$$EX = \sum_{x=0}^{K} x \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}} = \sum_{x=1}^{K} x \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}$$

Using the identities,

$$x\binom{M}{x} = M\binom{M-1}{x-1}$$
$$\binom{N}{K} = \frac{N}{K}\binom{N-1}{K-1},$$

we obtain

$$EX = \frac{KM}{N}\sum_{x=1}^{K}\frac{\binom{M-1}{x-1}\binom{N-M}{K-x}}{\binom{N-1}{K-1}} = \frac{KM}{N}.$$

A similar, but more lengthy, calculation will establish that

$$\text{Var}X = \frac{KM}{N}\left(\frac{(N-M)(N-K)}{N(N-1)}\right).$$

Example 3.1.1 *The hypergeometric distribution has application in acceptance sampling. Suppose a retailer buys goods in lots and each item can be either acceptable or defective. Let* $N = \#$ *of items in a lot, and* $M = \#$ *of defectives in a lot. Then we can calculate the probability that a sample of size* K *contains* x *defectives. To be specific, suppose that a lot of 25 machine parts is delivered, where a part is considered acceptable only if it passes tolerance. We sample 10 parts and find that none are defective (all are within tolerance). What is the probability of this event if there are 6 defectives in the lot of 25?*

Applying the hypergeometric distribution with $N = 25$, $M = 6$, *and* $k = 10$, *we have*

$$P(X = 0) = \frac{\binom{6}{0}\binom{19}{10}}{\binom{25}{10}} = 0.028,$$

showing that our observed event is quite unlikely if there 6 defectives in the lot.

3.1.3 Binomial Distribution

The binomial distribution is based on the idea of a *Bernoulli trial.* A Bernoulli trail is an experiment with two, and only two, possible outcomes. A random variable X has a Bernoulli(p) distribution if

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p, \end{cases}$$

where $0 \leq p \leq 1$. The value $X = 1$ is often termed a "success" and $X = 0$ is termed a "failure". The mean and variance of a Bernoulli(p) random variable are easily seen to be

$$EX = (1)(p) + (0)(1-p) = p$$

and

$$\text{Var}X = (1-p)^2 p + (0-p)^2(1-p) = p(1-p).$$

In a sequence of n identical, independent Bernoulli trials, each with success probability p, define the random variables $X_1, \ldots, X_n$ by

$$X_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p. \end{cases}$$

The random variable

$$Y = \sum_{i=1}^{n} X_i$$

has the binomial distribution. For this distribution,

$$EX = np, \qquad \mathrm{Var}X = np(1-p), \qquad M_X(t) = [pe^t + (1-p)]^n.$$

Theorem 3.1.1 *(Binomial theorem) For any real numbers x and y and integer $n \geq 0$,*

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}.$$

If we take $x = p$ and $y = 1-p$, we get

$$1 = (p + (1-p))^n = \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i}.$$

Example 3.1.2 *(Dice probabilities) Suppose we are interested in finding the probability of obtaining at least one 6 in four rolls of a fair die. This experiment can be modeled as a sequence of four Bernoulli trials with success probability* $p = \frac{1}{6}$*. Define*

$$X = \textit{total number of 6s in four rolls.}$$

Then $X \sim \textit{binomial}(4, \frac{1}{6})$ *and*

$$P(X > 0) = 1 - P(X = 0) = 1 - \binom{4}{0}(\frac{1}{6})^0(\frac{5}{6})^4 = 0.518.$$

Now we consider another game; throw a pair of dice 24 times and ask for the probability of at least one double 6. This, again, can be modeled by the binomial distribution with success probability p*, where*

$$p = P(\textit{roll a double 6}) = \frac{1}{36}/$$

Let Y =*number of double 6s in 24 rolls, then* $Y \sim \textit{binomial}(24, \frac{1}{36})$ *and*

$$P(Y > 0) = 1 - P(Y = 0) = 1 - \binom{24}{0}(\frac{1}{36})^0(\frac{35}{36})^{24} = 0.491.$$

This is the calculation originally done in the 18th century by Pascal at the request of the gambler de Mere, who thought both events had the same probability. He began to believe he was wrong when he started losing money on the second bet.

3.1.4 Poisson Distribution

The Poisson distribution has a single parameter λ, sometimes called the intensity parameter. A random variable X, taking values in the nonnegative integers, has a Poisson(λ) distribution if

$$P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots$$

The mean of X is

$$\begin{aligned} EX &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} \\ &= \lambda \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!} \\ &= \lambda \sum_{y=0}^{\infty} \frac{e^{-\lambda}\lambda^{y}}{y!} = \lambda \end{aligned}$$

A similar calculation will show that

$$\mathrm{Var} X = \lambda.$$

The mgf is

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

Example 3.1.3 *(Waiting time) As an example of a waiting-for-occurrence application, consider a telephone operator who, on the average, handles five calls every 3 minutes. What is the probability that there will be no calls in the next minute? At least two calls?*

If we let X =number of calls in a minute, then X has a Poisson distribution with $EX = \lambda = 5/3$. So

$$P(\textit{no calls in the next minute}) = P(X = 0)$$
$$= \frac{e^{-5/3}(\frac{5}{3})^0}{0!} = e^{-5/3} = 0.189$$

$$P(\textit{at least two calls in the next minute}) = P(X \geq 2)$$
$$= 1 - P(X = 0) - P(X = 1) = 1 - .189 - \frac{e^{-5/3}(5/3)^1}{1!} = 0.496.$$

Example 3.1.4 *(Poisson approximation) A typesetter, on the average, makes one error in every 500 words typeset. A typical page contains 300 words. What is the probability that there will be no more than two errors in five pages?*

If we assume that setting a word is a Bernoulli trial with success probability $p = \frac{1}{500}$, *then* X *=number of errors in five pages (1500 words) is binomial(1500,* $\frac{1}{500}$*). Thus,*

$$\begin{aligned} P(\textit{no more than two errors}) &= P(X \leq 2) \\ &= \sum_{x=0}^{2} \binom{1500}{x} (\frac{1}{500})^x (\frac{499}{500})^{1500-x} \\ &= .4230. \end{aligned}$$

If we use the Poisson approximation with $\lambda = 1500/500 = 3$, *we have*

$$P(X \leq 2) \approx e^{-3}(1 + 3 + \frac{3^2}{2}) = 0.4232.$$

3.1.5 Negative Binomial Distribution

In a sequence of independent Bernoulli(p) trials, let the random variable X denote the trial at which the r^{th} success occurs, where r is a fixed integer. Then

$$P(X = x|r,p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, \ldots, \tag{3.1}$$

and we say that X has a negative binomial(r, p) distribution.

The negative binomial distribution is sometimes defined in terms of the random variable Y =number of failures before rth success. This formulation is statistically equivalent to the one given above in terms of X =trial at which the rth success occurs, since $Y = X - r$. The alternative form of the negative binomial distribution is

$$P(Y = y) = \binom{r+y-1}{y} p^r 91 - p)^y, \quad y = 0, 1, \ldots.$$

The negative binomial distribution gets its name from the relationship

$$\binom{r+y-1}{y} = (-1)^y \binom{-r}{y} = (-1)^y \frac{(-r)(-r-1)\cdots(-r-y+1)}{(y)(y-1)\cdots(2)(1)},$$

which is the defining equation for binomial coefficient with negative integers.

$$
\begin{aligned}
EY &= \sum_{y=0}^{\infty} y \binom{r+y-1}{y} p^r (1-p)^y \\
&= \sum_{y=1}^{\infty} \frac{(r+y-1)!}{(y-1)!(r-1)!} p^r (1-p)^y \\
&= \sum_{y=1}^{\infty} \frac{r(1-p)}{p} \binom{r+y-1}{y-1} p^{r+1} (1-p)^{y-1} \\
&= \frac{r(1-p)}{p} \sum_{z=0}^{\infty} \binom{r+1+z-1}{z} p^{r+1} (1-p)^z \\
&= r\frac{1-p}{p}.
\end{aligned}
$$

A similar calculation will show

$$
\mathrm{Var} Y = \frac{r(1-p)}{p^2}.
$$

3.1.6 Geometric distribution

The geometric distribution is the simplest of the waiting time distributions and is a special case of the negative binomial distribution. Let $r = 1$ in (3.1) we have

$$P(X = x|p) = p(1-p)^{x-1}, \quad x = 1, 2, \ldots,$$

which defines the pmf of a geometric random variable X with success probability p.

X can be interpreted as the trial at which the first success occurs, so we are "waiting for a success". The mean and variance of X can be calculated by using the negative binomial formulas and by writing $X = Y + 1$ to obtain

$$EX = EY + 1 = \frac{1}{P} \quad \text{and} \quad \mathrm{Var}X = \frac{1-p}{p^2}.$$

The geometric distribution has an interesting property, known as the "memoryless" property. For integers $s > t$, it is the case that

$$P(X > s|X > t) = P(X > s - t), \tag{3.2}$$

that is, the geometric distribution "forgets" what has occurred. The probability of getting an additional $s - t$ failures, having already observed t failures, is the same as the probability of observing $s - t$ failures at the start of the sequence.

To establish (3.2), we first note that for any integer n,

$$P(X > n) = P(\text{no success in } n \text{ trials}) = (1-p)^n,$$

and hence,

$$\begin{aligned} P(X > s | X > t) &= \frac{P(X > s \text{ and } X > t)}{P(X > t)} = \frac{P(X > s)}{P(X > t)} \\ &= (1-p)^{s-t} = P(X > s - t). \end{aligned}$$

3.2 Continuous Distributions

3.2.1 Uniform Distribution

The continuous uniform distribution is defined by spreading mass uniformly over an interval $[a, b]$. Its pdf given by

$$f(x|a,b) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

It is easy to check that $\int_a^b f(x)dx = 1$ and

$$EX = \int_a^b \frac{x}{b-a} dx = \frac{b+a}{2}$$

$$\mathrm{Var}X = \int_a^b \frac{1}{b-a}(x - \frac{b+a}{2})^2 dx = \frac{(b-a)^2}{12}.$$

3.2.2 Gamma Distribution

The pdf of gamma(α, β) distribution is defined as

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/beta}, \quad 0 < x < \infty, \quad \alpha > 0, \quad \beta > 0.$$

The parameter α is known as the shape parameter, since it most influences the peakedness of the distribution, while the parameter β is called the scale parameter, since most of its influence is on the spread of the distribution.

$\Gamma(\alpha)$ is called the gamma function, it is defined as

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

It satisfies the following recursive formula

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha), \quad \alpha > 0.$$

$$\begin{aligned} EX &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x x^{\alpha-1} e^{-x/\beta} dx \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\alpha+1}\beta^{\alpha+1} = \alpha\beta \end{aligned}$$

In a manner analogous to the above, we can calculate EX^2 and then get

$$\text{Var}X = \alpha\beta^2.$$

The mgf of a gamma(α, β) distribution is

$$M_X(t) = (\frac{1}{1 - \beta t})^\alpha, \quad t < \frac{1}{\beta}.$$

Example 3.2.1 *(Gamma-Poisson relationship) There is an interesting relationship between the gamma and Poisson distributions. If X is a gamma(α, β) random variable, where α is an integer, then for any x,*

$$P(X \leq x) = P(Y \geq \alpha), \tag{3.3}$$

where $Y \sim$ Poisson(x/β).

There are a number of important special cases of the gamma distribution.

Chi-squared distribution If we set $\alpha = p/2$, where p is an integer, and $\beta = 2$, then the gamma pdf becomes

$$f(x|p) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{(p/2)-1} e^{-x/2}, \quad 0 < x < \infty,$$

which is the chi-squared pdf with p degrees of freedom.

exponential distribution If we set $\alpha = 1$, we have

$$f(x|\beta) = \frac{1}{\beta} e^{-x/\beta}, \quad 0 < x < \infty,$$

the exponential pdf with scale parameter β.

The exponential distribution can be used to model lifetimes, analogous to the use of the geometric distribution in the discrete case. In fact, the exponential distribution shares the "memoryless" property of the geometric.

$$\begin{aligned} P(X > s | X > t) &= \frac{P(X > s, X > t)}{P(X > t)} = \frac{P(X > s)}{P(X > t)} \\ &= \frac{\int_s^\infty \frac{1}{\beta} e^{-x/\beta} dx}{\int_t^\infty \frac{1}{\beta} e^{-x/\beta} dx} = \frac{e^{-s/\beta}}{e^{-t/\beta}} \\ &= e^{-(s-t)/\beta} = P(X > s - t). \end{aligned}$$

Weibull distribution f $X \sim$ exponential(β), then $Y = X^{1/\gamma}$ has a Weibull(γ, β) distribution,

$$f_Y(y|\gamma,\beta) = \frac{\gamma}{\beta} y^{\gamma-1} e^{-y^{\gamma}/\beta}, \quad 0 < y < \infty, \quad \gamma > 0, \quad ,\beta > 0.$$

The Weibull distribution plays an extremely important role in the analysis of failure time data.

3.2.3 Normal Distribution

The normal distribution has several advantages over the other distributions.

a. The normal distribution and distributions associated with it are very tractable and analytically.

b. The normal distribution has the familiar bell shape, whose symmetry makes it an appealing choice for many popular models.

c. There is the Central Limit Theorem, which shows that, under mild conditions, the normal distribution can be used to approximate a large variety of distributions in large samples.

The normal distribution has two parameters, usually denoted by μ and σ^2, which are its mean and variance. The pdf is

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}, \quad -\infty < x < \infty.$$

If $X \sim N(\mu,\sigma^2)$, then the random variable $Z = (X-\mu)/\sigma$ has a $N(0,1)$ distribution, also known as the standard normal.

If $Z \sim N(0,1)$,

$$EZ = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2} dz = 0,$$

and so, if $X \sim N(\mu, \sigma^2)$,

$$EX = E(\mu + \sigma Z) = \mu + \sigma EZ = \mu.$$

Similarly, we have that Var$Z = 1$ and Var$X = \sigma^2$.

To show

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz = 1.$$

We only need to show

$$\int_0^{\infty} e^{-z^2/2} dz = \sqrt{\frac{\pi}{2}}.$$

Since

$$\begin{aligned}\Big(\int_0^{\infty} e^{-z^2/2} dz\Big)^2 &= \Big(\int_0^{\infty} e^{-t^2/2} dt\Big)\Big(\int_0^{\infty} e^{-u^2/2} du\Big) \\ &= \int_0^{\infty} \int_0^{\infty} e^{-(t^2+u^2)/2} dt du.\end{aligned}$$

Now we convert to polar coordinates. Define

$$t = r\cos\theta, \quad u = r\sin\theta.$$

Then $t^2 + u^2 = r^2$ and $dtdu = rd\theta dr$ and the limits of integration become $0 < r < \infty$, $0 < \theta < \pi/2$. We now have

$$\begin{aligned}\int_0^{\infty} \int_0^{\infty} e^{-(t^2+u^2)/2} dt du &= \int_0^{\infty} \int_0^{\infty} r e^{-r^2/2} d\theta dr \\ &= \frac{\pi}{2} \int_0^{\infty} r e^{-r^2/2} dr = \frac{\pi}{2}.\end{aligned}$$

The probability content within 1, 2 or 3 standard deviations of the mean is

$$P(|X-\mu| \le \sigma) = P(|Z| \le 1) = .6826,$$

$$P(|X-\mu| \le 2\sigma) = P(|Z| \le 2) = .9544,$$

$$P(|X-\mu| \le 3\sigma) = P(|Z| \le 3) = .9974,$$

where $X \sim N(\mu, \sigma^2)$ and $Z \sim N(0,1)$.

Among the many uses of the normal distribution, an important one is its use as an approximation to other distributions. For example, if $X \sim \text{binomial}(n,p)$, then $EX = np$ and $\text{Var}X = np(1-p)$, and under suitable conditions, the distribution of X can be approximated by that of a normal random variable with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$. The suitable conditions are that n should be large and p should not be extreme (near 0 or 1). We want n large so that there are enough values of X to make an approximation by a continuous distribution reasonable, and p should be "in the middle" so the binomial is nearly symmetric, as is the normal. A conservative rule to follow is that the approximation will be good if $\min(np, n(1-p)) \ge 5$.

3.2.4 Beta Distribution

The beta(α, β) pdf is

$$f(x|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1, \quad \alpha > 0, \quad \beta > 0,$$

where $B(\alpha, \beta)$ denotes the beta function,

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

For $n > -\alpha$, we have

$$\begin{aligned} EX^n &= \frac{1}{B(\alpha,\beta)} \int_0^1 x^n x^{\alpha-1}(1-x)^{\beta-1}dx \\ &= \frac{B(\alpha+n,\beta)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+n)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)\Gamma(\alpha)}. \end{aligned}$$

Then mean and variance are

$$EX = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad \mathrm{Var}X = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

3.2.5 Cauchy Distribution

The Cauchy distribution is a symmetric, bell-shaped distribution on $(-\infty, \infty)$ with pdf

$$f(x|\theta) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}, \quad -\infty < x < \infty, \quad -\infty < \theta < \infty.$$

The mean of Cauchy distribution does not exist, that is,

$$E|X| = \int_{-\infty}^{\infty} \frac{1}{\pi}\frac{|x|}{1+(x-\theta)^2}dx = \infty.$$

Since $E|X| = \infty$, it follows that no moments of the Cauchy distribution exist. In particular, the mgf does not exist.

3.2.6 Lognormal Distribution

If X is a random variable whose logarithm is normally distributed, then X has a lognormal distribution. The pdf of X can be obtained by straightforward transformation of the normal pdf, yielding

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{x}e^{-(\log x - \mu)^2/(2\sigma^2)}, \quad 0 < x < \infty, \quad -\infty < \mu < \infty, \sigma >$$

for the lognormal pdf.

$$EX = Ee^{\log X} = Ee^{Y} = e^{\mu + (\sigma^2/2)}.$$

$$\text{Var}X = e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}.$$

3.2.7 Double Exponential Distribution

The double exponential distribution is formed by reflecting the exponential distribution around its mean. The pdf is given by

$$f(x|\mu, \sigma) = \frac{1}{2\sigma}e^{-|x-\mu|/\sigma}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0.$$

It is straightforward to calculate

$$EX = \mu \quad \text{and} \quad \text{Var}X = 2\sigma^2.$$

3.3 Exponential Families

A family of pdfs or pmfs is called an exponential family if it can be expressed as

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta}) \exp\big(\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(x)\big). \tag{3.4}$$

Here $h(x) \geq 0$ and $t_1(x), \ldots, t_k(x)$ are real-valued functions of the observation x (they cannot depend on $\boldsymbol{\theta}$), and $c(\boldsymbol{\theta}) \geq 0$ and $w_1(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta})$ are real-valued functions of the possibly vector-valued parameter $\boldsymbol{\theta}$ (they cannot depend on x). Many common families introduced in the previous section are exponential families. They include the continuous families—normal, gamma, and beta, and the discrete families—binomial, Poisson, and negative binomial.

Example 3.3.1 *(Binomial exponential family)*

Let n be a positive integer and consider the binomial(n, p) family with $0 < p < 1$. Then the pmf for this family, for $x = 0, 1, \ldots, n$ and $0 < p < 1$, is

$$\begin{aligned} f(x|p) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= \binom{n}{x} (1-p)^n \Big(\frac{p}{1-p}\Big)^x \\ &= \binom{n}{x} (1-p)^n \exp\Big(\log\Big(\frac{p}{1-p}\Big)x\Big). \end{aligned}$$

Define

$$h(x) = \begin{cases} \binom{n}{x} & x = 0, 1, \ldots, n \\ 0 & \textit{otherwise}, \end{cases}$$

$$c(p) = (1-p)^n, \quad 0 < p < 1, \quad w_1(p) = \log\Big(\frac{p}{1-p}\Big), \quad 0 < p < 1,$$

and

$$t_1(x) = x.$$

Then we have

$$f(x|p) = h(x)c(p)\exp\{w_1(p)t_1(x)\}.$$

Example 3.3.2 *(Normal exponential family) Let* $f(x|\mu, \sigma^2)$ *be the* $N(\mu, \sigma^2)$ *family of pdfs, where* $\boldsymbol{\theta} = (\mu, \sigma^2)$, $-\infty < \mu < \infty$, $\sigma > 0$. *Then*

$$\begin{aligned} f(x|\mu, \sigma^2) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\big(- \frac{(x-\mu)^2}{2\sigma^2}\big) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\big(- \frac{\mu^2}{2\sigma^2}\big) \exp\big(- \frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}\big). \end{aligned}$$

Define

$$h(x) = 1 \textit{ for all } x;$$

$$c(\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\big(- \frac{\mu^2}{2\sigma^2}\big), -\infty < \mu < \infty, \sigma > 0;$$

$$w_1(\boldsymbol{\theta}) = \frac{1}{\sigma^2}, \quad \sigma > 0; \qquad w_2(\boldsymbol{\theta}) = \frac{\mu}{\sigma^2}, \quad \sigma > 0;$$

$$t_1(x) = -x^2/2; \quad \textit{and} \quad t_2(x) = x.$$

Then

$$f(x|\mu, \sigma^2) = h(x)c(\mu, \sigma) \exp[w_1(\mu, \sigma)t_1(x) + w_2(\mu, \sigma)t_2(x)].$$

Definition 3.3.1 *The indicator function of a set A, most often denoted by $I_A(x)$, is the function*

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}$$

An alternative notation is $I(x \in A)$.

In general, the set of x values for which $f(x|\boldsymbol{\theta} > 0$ cannot depend on $\boldsymbol{\theta}$ in an exponential family. For example, the set of pdfs given by

$$f(x|\theta) = \theta^{-1} \exp(1 - x/\theta), \quad 0 < \theta < x < \infty,$$

is not an exponential family. We have

$$f(x|\theta) = \theta^{-1} \exp(1 - x/\theta) I_{[\theta,\infty)}(x).$$

The indicator function can not be incorporated into any of the functions of (3.4) since it is not a function of x alone, not a function of θ alone, and cannot be expressed as an exponential.

Definition 3.3.2 *A curved exponential family is a family of densities of the form (3.4) for which the dimension of the vector* $\boldsymbol{\theta}$ *is equal to* $d < k$*. If* $d = k$*, the family is a full exponential family.*

Example 3.3.3 *(A curved exponential family) The normal family of Example 3.3.2 is a full exponential family. However, if we assume that* $\sigma^2 = \mu^2$*, the family becomes curved. We have*

$$\begin{aligned} f(x|\mu) &= \frac{1}{\sqrt{2\pi\mu^2}} \exp\big(-\frac{(x-\mu)^2}{2\mu^2}\big) \\ &= \frac{1}{\sqrt{2\pi\mu^2}} \exp(-\frac{1}{2}) \exp(-\frac{x^2}{2\mu^2} + \frac{x}{\mu}). \end{aligned}$$

The full exponential family would have parameter space $(\mu, \sigma^2) = (-\infty, \infty) \times (0, \infty)$*, while the parameter space of the curved family* $(\mu, \sigma^2) = (\mu, \mu^2)$ *is a parabola.*

3.4 Location and Scale Families

Theorem 3.4.1 *let $f(x)$ be any pdf and let μ and $\sigma > 0$ be any given constants. Then the function*

$$g(x|\mu,\sigma) = \frac{1}{\sigma} f(\frac{x-\mu}{\sigma})$$

is a pdf.

PROOF: Since $f(x) \geq 0$ for all values of x. So, $\frac{1}{\sigma} f(\frac{x-\mu}{\sigma}) \geq 0$ for all values of x, μ and σ. Next,

$$\int_{-\infty}^{\infty} \frac{1}{\sigma} f(\frac{x-\mu}{\sigma}) dx = \int_{-\infty}^{\infty} f(y) dy = 1.$$

□

Definition 3.4.1 *Let $f(x)$ be any pdf. Then the family of pdfs $f(x-\mu)$, indexed by the parameter μ, $-\infty < \mu < \infty$, is called the location family with standard pdf $f(x)$ and μ is called the location parameter for the family.*

The location parameter μ simply shifts the pdf $f(x)$ so that the shape of the graph is unchanged but the point on the graph that was above $x = 0$ for $f(x)$ is above $x = \mu$ for $f(x - \mu)$.

Example 3.4.1 *(Exponential location family) Let $f(x) = e^{-x}$, $x \geq 0$, and $f(x) = 0$, $x < 0$. To form a location family we replace*

x with $x - \mu$ to obtain

$$f(x|\mu) = \begin{cases} e^{-(x-\mu)} & x - \mu \geq 0 \\ 0 & x - \mu < 0. \end{cases}$$

That is,

$$f(x|\mu) = \begin{cases} e^{-(x-\mu)} & x \geq \mu \\ 0 & x < \mu. \end{cases}$$

In this type of model, where μ denotes a bound on the range of X, μ is sometimes called a threshold parameter.

Definition 3.4.2 *Let $f(x)$ be any pdf. Then for any $\sigma > 0$, the family of pdfs $(1/\sigma)f(x/\sigma)$, indexed by the parameter σ, is called the scale family with standard pdf $f(x)$ and σ is called the scale parameter of the family.*

The effect of introducing the scale parameter σ is either to stretch $(\sigma > 1)$ or to contract $(\sigma < 1)$ the graph of $f(x)$ while still maintaining the same basic shape of the graph.

Definition 3.4.3 *Let $f(x)$ be any pdf. Then for any μ, $-\infty < \mu < \infty$, and any $\sigma > 0$, the family of pdfs $(1/\sigma)f((x-\mu)/\sigma)$, indexed by the parameter (μ, σ), is called the location-scale family with standard pdf $f(x)$; μ is called the location parameter and σ is called the scale parameter.*

The effect of introducing both the location and scale parameters is to stretch ($\sigma > 1$) or contract ($\sigma < 1$) the graph with the scale parameter and then shift the graph so that the point that was above 0 is now above μ. The normal and double exponential families are examples of location-scale families.

Theorem 3.4.2 *Let $f(\cdot)$ be any pdf. Let μ be any real number, and let σ be any positive real number. Then X is a random variable with pdf $(1/\sigma)f((x-\mu)/\sigma)$ if and only if there exists a random variable Z with pdf $f(z)$ and $X = \sigma Z + \mu$.*

Theorem 3.4.3 *Let Z be a random variable with pdf $f(z)$. Suppose EZ and $VarZ$ exist. If X is a random variable with pdf $(1/\sigma)f((x-\mu)/\sigma)$, then*

$$EX = \sigma EZ + \mu \quad \textit{and} \quad VarX = \sigma^2 VarZ.$$

In particular, if $EZ = 0$ and $VarZ = 1$, then $EX = \mu$ and $VarX = \sigma^2$.

3.5 Inequalities and Identities

Theorem 3.5.1 *(Chebychev's Inequality) Let X be a random variable and let $g(x)$ be a nonnegative function. Then, for any $r > 0$,*

$$P(g(X) \geq r) \leq \frac{Eg(X)}{r}.$$

PROOF:

$$\begin{aligned} Eg(X) &= \int_{-\infty}^{\infty} g(x) f_X(x) dx \\ &\geq \int_{\{x: g(x) \geq r\}} g(x) f_X(x) dx \quad (g \text{ is nonnegative}) \\ &\geq r \int_{\{x: g(x) \geq r\}} f_X(x) dx \\ &= rP(g(X) \geq r). \end{aligned}$$

Rearranging now produces the desired inequality. □

Example 3.5.1 *(Illustrating Chebychev) let* $g(x) = (x-\mu)^2/\sigma^2$*, where* $\mu = EX$ *and* $\sigma^2 = VarX$*. For convenience write* $r = t^2$*. Then*

$$P(\frac{(X-\mu)^2}{\sigma^2} \geq t^2) \leq \frac{1}{t^2} E\frac{(X-\mu)^2}{\sigma^2} = \frac{1}{t^2}.$$

Thus,

$$P(|X-\mu| \geq t\sigma) \leq \frac{1}{t^2}.$$

For example, taking $t = 2$*, we get*

$$P(|X-\mu| \geq 2\sigma) \leq \frac{1}{2^2} = 0.25.$$

Example 3.5.2 *(A normal probability inequality) If* Z *is standard normal, then*

$$P(|Z| \geq t) \leq \sqrt{\frac{2}{\pi}} \frac{e^{-t^2/2}}{t}, \quad \textit{for all } t > 0.$$

Write

$$\begin{aligned} P(Z \geq t) &= \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-x^2/2} dx \\ &\leq \frac{1}{\sqrt{2\pi}} \int_t^\infty \frac{x}{t} e^{-x^2/2} dx \quad (\textit{since } x/t > 1) \\ &= \frac{1}{\sqrt{2\pi}} \frac{e^{-t^2/2}}{t} \end{aligned}$$

and use the fact that $P(|Z| \geq t) = 2P(Z \geq t)$*.*

Theorem 3.5.2 *Let $X_{\alpha,\beta}$ denote a gamma(α, β) random variable with pdf $f(x|\alpha, \beta)$, where $\alpha > 1$. Then for any constants a and b,*

$$P(a < X_{\alpha,\beta} < b) = \beta f(a|\alpha, \beta) - f(b|\alpha, \beta) + P(a < X_{\alpha-1,\beta} < b).$$

Lemma 3.5.1 *(Stein's Lemma) Let $X \sim N(\theta, \sigma^2)$, and let g be a differentiable function satisfying $E|g'(X)| < \infty$. Then*

$$E[g(X)(X - \theta)] = \sigma^2 Eg'(X).$$

PROOF: The left-hand side is

$$E[g(X)(X - \theta)] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} g(x)(x - \theta)e^{-(x-\theta)^2/(2\sigma^2)} dx.$$

Using integration by parts with $u = g(x)$ and $dv = (x-\theta)e^{-(x-\theta)^2/(2\sigma^2)}dx$ to get

$$E[g(X)(X-\theta)] = \frac{1}{\sqrt{2\pi}\sigma} \left[-\sigma^2 g(x)e^{-(x-\theta)^2/(2\sigma^2)}\Big|_{-\infty}^{\infty} + \sigma^2 \int_{-\infty}^{\infty} g'(x)e^{-(x-\theta)^2/}\right.$$

The condition on g' is enough to ensure that the first term is 0 and what remains on the right-hand side is $\sigma^2 Eg'(X)$. $\square$

Example 3.5.3 *(Higher-order normal moments) Stein's lemma makes calculation of higher-order moments quite easy/ For example, if* $X \sim N(\theta, \sigma^2)$*, then*

$$\begin{aligned} EX^3 &= EX^2(X - \theta + \theta) = EX^2(X - \theta) + \theta EX^2 \\ &= 2\sigma^2 EX + \theta EX^2 = 2\sigma^2\theta + \theta(\sigma^2 + \theta^2) \\ &= 3\theta\sigma^2 + \theta^3. \end{aligned}$$

Theorem 3.5.3 *Let* χ^2_p *denote a chi-squared random variable with* p *degrees of freedom. For any function* $h(x)$*,*

$$Eh(\chi^2_p) = pE\Big(\frac{h(\chi^2_{p+2})}{\chi^2_{p+2}}\Big)$$

provided the expectations exist.

Some moment calculations are very easy with Theorem 3.5.3. For example, the mean of a χ^2_p is

$$E\chi^2_p = pE\Big(\frac{\chi^2_p}{\chi^2_p}\Big) = pE(1) = p,$$

and the second moment is

$$E(\chi^2_p)^2 = pE\Big(\frac{(\chi^2_p)^2}{\chi^2_p}\Big) = pE(\chi^2_p) = p(p+2).$$

So $\mathrm{Var}\chi^2_p = p(p+2) - p^2 = 2p$.

Theorem 3.5.4 *(Hwang) Let $g(x)$ be a function with $-\infty < Eg(X) < \infty$ and $-\infty < g(-1) < \infty$. Then:*

a. If $X \sim Poisson(\lambda)$,

$$E(\lambda g(X)) = E(Xg(X-1)).$$

b. If $X \sim$ negative binomial(r,p),

$$E((1-p)g(X)) = E\Big(\frac{X}{r+X-1}g(X-1)\Big).$$

Example 3.5.4 *(Higher-order Poisson moments) For* $X \sim$ *Poisson*(λ)*, take* $g(x) = x^2$ *and use Theorem 3.5.4:*

$$E(\lambda X^2) = E(X(X-1)^2) = E(X^3 - 2X^2 + X).$$

Therefore, the third moment of a Poisson(λ) *is*

$$\begin{aligned} EX^3 &= \lambda EX^2 = 2EX^2 - EX \\ &= \lambda(\lambda + \lambda^2) + 2(\lambda + \lambda^2) - \lambda = \lambda^3 + 3\lambda^2 + \lambda. \end{aligned}$$

For the negative binomial, the mean can be calculated by taking $g(x) = r + x$,

$$E((1-p)(r+X)) = E\Big(\frac{X}{r+X-1}(r+X-1)\Big) = EX,$$

so, rearranging, we get

$$EX = \frac{r(1-p)}{p}.$$