P8109-STATISTICAL INFERENCE

TABLE OF CONTENTS

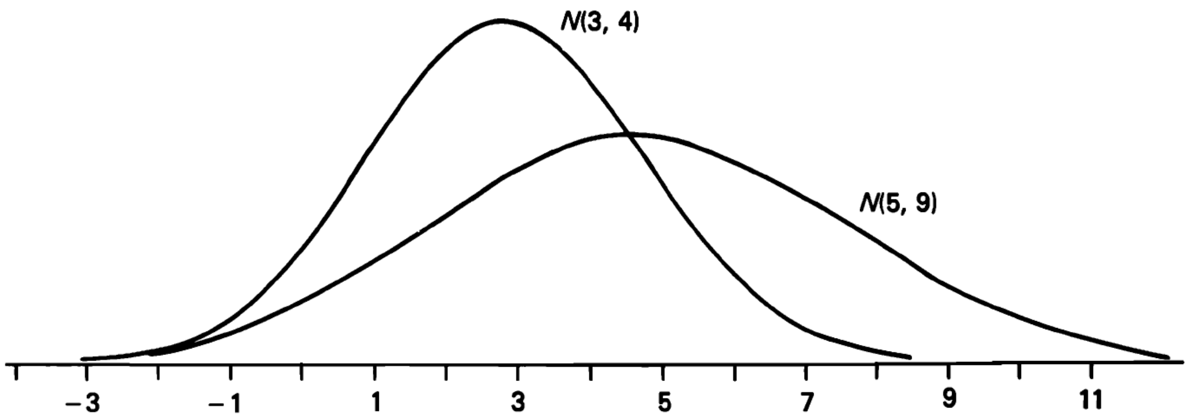
1 NORMAL AND ASSOCIATED DISTRIBUTIONS

1.1 THE NORMAL DISTRIBUTION

If a random variable (r.v.) X has probability density function (p.d.f.)

$$f_X(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \qquad -\infty < x < \infty, -\infty < \mu < \infty, 0 < \sigma \tag{1-1}$$

then X is said to have a normal distribution with mean μ and variance σ^2, i.e. $X \sim N(\mu, \sigma^2)$. μ is called a location parameter while σ is called a scale parameter.



By integration, it can be shown that if $X \sim N(\mu,\ \sigma^2)$ then the moment-generating function (m.g.f.) of X is

$$M_X(t) = \mathrm{E}e^{tX} = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right) \tag{1-2}$$

Furthermore, if we define $Z = (X-\mu)/\sigma$, then $Z \sim N(0,\ 1)$ is the *standard normal* distribution:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right) \qquad -\infty < z < \infty$$

Example 1.1A

Show that $\int_{-\infty}^{\infty} f_Z(z)dz = 1$.

Solution

Let $I = \int_{-\infty}^{\infty} e^{-z^2/2}dz = \int_{-\infty}^{\infty} e^{-x^2/2}dx = \int_{-\infty}^{\infty} e^{-y^2/2}dy$. Then

$$
\begin{aligned}
I^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2/2} dx\right)\left(\int_{-\infty}^{\infty} e^{-y^2/2} dy\right) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dxdy \\
&= \int_0^{2\pi}\int_0^{\infty} e^{-(x^2+y^2)/2} r dr d\theta && \begin{bmatrix}\text{by changing to polar coordinates, i.e.} \\ x = r\cos\theta, y = r\sin\theta\end{bmatrix} \\
&= \int_0^{2\pi}\left(\int_0^{\infty} re^{-r^2/2} dr\right) d\theta \\
&= \int_0^{2\pi}\left(\int_0^{\infty} e^{-u} du\right) d\theta && \left[\text{by substituting } u = r^2/2\right] \\
&= \int_0^{2\pi} d\theta \\
&= 2\pi
\end{aligned}
$$

Hence,

$$\int_{-\infty}^{\infty} f_Z(z)dz = \frac{1}{\sqrt{2\pi}} I = \frac{\sqrt{2\pi}}{\sqrt{2\pi}} = 1\,.$$

An important result concerns the linear combination of r.v.'s:

<u>Theorem 1.1A</u>: If $X \sim N(\mu, \sigma^2)$, then $aX + b \sim N\left(a\mu + b, a^2\sigma^2\right)$.

<u>Proof</u>. Using the result for the m.g.f. of X in Eq. (1-2), the m.g.f. of $aX + b$ is

$$
\begin{aligned}
\mathrm{E}e^{t(aX+b)} &= e^{bt}\mathrm{E}e^{atX} \\
&= e^{bt}e^{\mu(at)+\frac{1}{2}\sigma^2(at)^2} \\
&= e^{(\mu a+b)t+\frac{1}{2}(a\sigma)^2t^2}
\end{aligned}
$$

Hence, $aX + b \sim N\left(a\mu + b, a^2\sigma^2\right)$ ■

More generally:

Theorem 1.1B If $X_1,\ldots,X_n$ are independent r.v.s with each $X_i \sim N(\mu_i,\ \sigma_i^2)$ and $W = \sum_i a_i X_i + b$, then W is normal with

$$\mathrm{E}W = \sum_i a_i \mu_i + b \quad \text{and} \quad \mathrm{var}\,W = \sum_i a_i^2 \sigma_i^2$$

Proof. As in Theorem 1.1A■

Remark: The converse of the above is not true: if a linear combinations of independent r.v.'s is normally distributed, then it does not follow that each r.v. is normally distributed. However, Bernstein proved in 1941 that if *X* and *Y* are independent and identical distributed (i.i.d.) r.v.'s such that *X* + *Y* and *X* – *Y* are independent, then *X* and *Y* are each normally distributed.

1.2 THE χ^2-, t-, AND F-DISTRIBUTIONS

(a) Let $Z_1,\ldots,Z_k$ be independent N(0, 1) random variables. Then

$$\sum_{i=1}^{k} Z_i^2 \sim \chi_k^2, \tag{1-3}$$

where χ_k^2 is the chi-squared distribution with *k* degrees of freedom (d.f.). Recall that, in general, the χ_k^2 -distribution is a special case of the gamma(*α*, *β*) family. If *U*~gamma (*α*, *β*), then

$$\begin{aligned} f_U(u;\alpha,\beta) &= \frac{u^{\alpha-1}e^{-u/\beta}}{\beta^{\alpha}\Gamma(\alpha)}, \quad u>0,\alpha>0,\beta>0 \\ &= \frac{u^{\alpha-1}e^{-u/\beta}}{\beta^{\alpha}\Gamma(\alpha)} I(u>0), \quad \alpha>0,\beta>0 \end{aligned},$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$ ($\alpha > 0$) is the gamma function, and $I(E) = 1$ if the event *E* is true and $I(E) = 0$ otherwise. It can be shown that $\mathrm{E}U = \alpha\beta$ and $\mathrm{var}\,U = \alpha\beta^2$. Note also that $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$ for all real α and $\Gamma(\alpha) = (\alpha-1)!$ for $\alpha = 1, 2, \ldots$

If *U*~gamma (*k*/2, 2), then $U \sim \chi_k^2$. The mean and variance of *U* are *k* and 2*k*, respectively. The m.g.f. of U is $(1-2t)^{-k/2}$.

If U~gamma (1, β), then U has an exponential distribution with parameter $\beta > 0$. The mean is β, the variance is β^2, and the m.g.f. is $(1-\beta t)^{-1}$ for $t < 1/\beta$.

(b) If $Z \sim N(0, 1)$, $V \sim \chi_k^2$, and Z and V are independent then

$$\frac{Z}{\sqrt{V/k}} \sim t_k \tag{1-4}$$

where t_k is the t-distribution with k d.f. Recall that if $T \sim t_k$ then

$$f_T(t;k) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}}\left(1+\frac{t^2}{k}\right)^{-(k+1)/2}, \qquad -\infty < t < \infty .$$

Moreover, $\mathrm{E}T = 0$ if $k > 1$ and $\operatorname{var} T = k/(k-2)$ for $k > 2$. For large d.f. the t-distribution converges to the standard normal, i.e. $T \xrightarrow{D} Z \sim N(0, 1)$. The m.g.f. of a t-distribution does not exist.

(c) If $U \sim \chi_m^2$, $V \sim \chi_n^2$, and U and V are independent then

$$\frac{U/m}{V/n} \sim F_{m,n}, \tag{1-5}$$

where $F_{m,n}$ is the F-distribution with m numerator d.f. and n denominator d.f. Recall that if $X \sim F_{m,n}$ then

$$\begin{aligned} f_X(x;m,n) &= \frac{\Gamma\left(\frac{m+n}{2}\right)(m/n)^{m/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{x^{(m/2)-1}}{\left(1+\frac{mx}{n}\right)^{(m+n)/2}}, \qquad x > 0 \\ &= \frac{\Gamma\left(\frac{m+n}{2}\right)(m/n)^{m/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{x^{(m/2)-1}}{\left(1+\frac{mx}{n}\right)^{(m+n)/2}} I(x > 0) \end{aligned}$$

Some results associated with the F- distribution are:

(i) If $T \sim t_k$, then $T^2 \sim F_{1,k}$;

(ii) If $X \sim F_{m,n}$, then $1/X \sim F_{n,m}$;

(iii) If $X \sim F_{m,n}$, $\mathrm{E}X = n/(n-2)$ for $n > 2$.

1.3 THE CENTRAL-LIMIT THEOREM

There are several versions of the Central Limit Theorem (CLT), each which different degrees of generality. Here we shall state and prove a less general (and more common) version known as the Lindeberg-Lévy CLT.

Theorem 1.3A (Lindeberg-Lévy CLT) If $X_1, X_2, ..., X_n$ are i.i.d. r.v.'s with $\mathrm{E}X_i = \mu$ and $\mathrm{var}\, X_i = \sigma^2$. Then the sum $S_n = X_1 + X_2 + ... + X_n$ is asymptotical normal with mean $n\mu$ and variance $n\sigma^2$, i.e.

$$S_n \dot{\underset{\cdot}{\sim}} N\left(n\mu, n\sigma^2\right) \quad \text{or} \quad \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{D} Z \sim N(0,\ 1) \tag{1-6}$$

The above result can also be written in terms of the sample mean $\bar{X}_n = S_n / n$:

$$\bar{X}_n \dot{\underset{\cdot}{\sim}} N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{or} \quad \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} Z \sim N(0,\ 1) \tag{1-7}$$

■

Example 1.3A

Apply the Lindeberg-Lévy CLT to $\sum_{i=1}^{k} Z_i^2$ in Eq. (1-3) and deduce it asymptotic distribution.

Solution

Since $Z_i \sim N(0,\ 1)$, $\mathrm{E}Z_i = 0$ and $\mathrm{var}\, Z_i = 1$. To apply the CLT, we need the mean and variance of each Z_i^2. We have $\mathrm{E}Z_i^2 = \mathrm{E}\chi_1^2 = 1$, and $\mathrm{var}\, Z_i^2 = \mathrm{var}\, \chi_1^2 = 2$. Therefore, by the Lindeberg-Lévy CLT,

$\sum_{i=1}^{k} Z_i^2 \dot{\underset{\cdot}{\sim}} N(k, 2k)$, i.e. for large d.f. the χ_k^2-distribution approaches the $N(k, 2k)$ distribution.

<u>Proof Theorem 1.3A (Lindeberg-Lévy CLT)</u> Let $Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Then the m.g.f. of Z_n is

$$
\begin{aligned}
M_{Z_n}(t) &= \mathrm{E}e^{tZ_n} \\
&= \mathrm{E}\exp\left\{t\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right)\right\} \\
&= \mathrm{E}\exp\left(t\frac{X_1-\mu}{\sigma\sqrt{n}} + \ldots + t\frac{X_n-\mu}{\sigma\sqrt{n}}\right) \\
&= \left[\mathrm{E}\exp\left(t\frac{X_i-\mu}{\sigma\sqrt{n}}\right)\right]^n \\
&= \left[\mathrm{E}\left\{1 + t\frac{X_i-\mu}{\sigma\sqrt{n}} + \frac{t^2}{2}\left(\frac{X_i-\mu}{\sigma\sqrt{n}}\right)^2 + O\left(t^3\right)\right\}\right]^n \quad [\text{since } e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots] \\
&= \left[\left\{1 + t(0) + \frac{t^2}{2\sigma^2 n}\operatorname{var} X_i + O\left(t^3\right)\right\}\right]^n \\
&= \left\{1 + \frac{t^2}{2n} + O\left(t^3\right)\right\}^n \\
&\to e^{t^2/2} \quad \text{as} \quad n \to \infty
\end{aligned}
$$

But this is the m.g.f. of a N(0, 1) r.v. Hence the Lindeberg-Lévy CLT theorem is proved. ■

1.4 SAMPLING FROM A NORMAL DISTRIBUTION

Theorem 1.4A. Let $X_1, X_2, ..., X_n$ be i.i.d. $N(\mu, \sigma^2)$ r.v.'s. Define the sample mean and sample variance respectively as follows:

$$
\bar{X} = \frac{\sum_i X_i}{n}, \qquad S^2 = \frac{\sum_i \left(X_i - \bar{X}\right)^2}{n-1}
$$

Then:

(a) $\bar{X}$ and S^2 are independent; (1-8)

(b) $\bar{X} \sim N(\mu,\ \sigma^2 / n)$; (1-9)

(c) $(n-1)S^2 / \sigma^2 \sim \chi^2_{n-1}$. (1-10)

<u>Proof</u> (a) We use the *Helmert* transformation:

$$
\begin{aligned}
U_1 &= \left(X_1 - X_2\right) / \sqrt{2}, \\
U_2 &= \left(X_1 + X_2 - 2X_3\right) / \sqrt{6}, \\
U_3 &= \left(X_1 + X_2 + X_3 - 3X_4\right) / \sqrt{12}, \\
&\ldots \\
U_{n-1} &= \left\{X_1 + X_2 + \ldots + X_{n-1} - \left(n-1\right) X_n\right\} / \sqrt{n\left(n-1\right)}, \\
U_n &= \left(X_1 + X_2 + \ldots + X_n\right) / \sqrt{n}
\end{aligned}
$$

If we write $U_i = \sum_{j=1}^{n} a_{ij} X_j \quad (i = 1, \ldots, n)$, then it is seen that

$$\sum_{i=1}^{n} a_{ij} a_{ik} = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases}$$

This implies that the transformation of the X_j's to the U_i's are orthogonal. Using the theorem that an orthogonal transformation of independent r.v.s each normally distributed with variance σ^2 produces new r.v.s which are also independent and each normally distributed with variance σ^2, we see that the U_i's are independent and each normally distributed with variance σ^2. Furthermore,

$$
\begin{aligned}
\sum_{i=1}^{n} U_i^2 &= \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ij} a_{ik} X_j X_k \\
&= \sum_{j=1}^{n} \sum_{k=1}^{n} X_j X_k \left(\sum_{i=1}^{n} a_{ij} a_{ik} \right) \\
&= \sum_{j=1}^{n} X_j^2,
\end{aligned}
$$

so that

$$U_1^2 + U_2^2 + \ldots + U_{n-1}^2 = \sum_{i=1}^{n} X_i^2 - U_n^2$$
$$= \sum_{i=1}^{n} X_i^2 - n\overline{X}^2$$
$$= \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2$$

Since U_n is independent of $U_1^2 + U_2^2 + \ldots + U_{n-1}^2$, we obtain that $\overline{X} = U_n / \sqrt{n}$ is independent of $U_1^2 + U_2^2 + \ldots + U_{n-1}^2 = \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2$ and hence of S^2.

(b) We write $\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$ and apply the result in Theorem 1.1B. Therefore, $\bar{X}$ is normal with

$$\mathrm{E}\bar{X} = \sum_i \frac{1}{n}(\mu) = \mu,$$

$$\operatorname{var} \bar{X} = \sum_i \frac{1}{n^2}\sigma^2 = \frac{\sigma^2}{n}.$$

(c) Define $A = \sum_{i=1}^{n}(X_i - \mu)^2 / \sigma^2$ and $C = (\bar{X} - \mu)^2 / (\sigma^2 / n)$. From the result in (1-3), $A \sim \chi_n^2$ and $C \sim \chi_1^2$.

Now,

$$\underbrace{\frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma^2}}_{A} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X} + \bar{X} - \mu\right)^2}{\sigma^2} = \underbrace{\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{\sigma^2}}_{B} + \underbrace{\frac{\sum_{i=1}^{n}\left(\bar{X} - \mu\right)^2}{\sigma^2}}_{C}$$

since $\sum_{i=1}^{n}(X_i - \bar{X})(\bar{X} - \mu) = 0$. By (a) above, *B* and *C* are independent, so that

$$M_A(t) = M_B(t) M_C(t) \qquad \Rightarrow \qquad M_B(t) = \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}} = (1-2t)^{-(n-1)/2}$$

which is the m.g.f. of the χ_{n-1}^2-distribution. Hence

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i==1}^{n}\left(X_i - \bar{X}\right)^2}{\sigma^2} \sim \chi_{n-1}^2. \blacksquare$$

Remarks: (i) The independence in (a) above <u>does not</u> hold when sampling from a non-normal distribution. It characterizes the normal distribution.

(ii) The result in (b) is <u>not</u> a consequence of the CLT.

Two important corollaries to Theorem 1.4A above are that:

(1) $\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$ (1-11)

(2) $\mathrm{E}S^2 = \sigma^2, \mathrm{var}\, S^2 = \dfrac{2\sigma^4}{(n-1)}$ (1-12)

<u>Proof</u> (1) From the result in (1-4) and the result (c) in Theorem 1.4A ,

$$T = \frac{Z}{\sqrt{\dfrac{(n-1)S^2/\sigma^2}{n-1}}} \sim t_{n-1}$$

$$\therefore \quad \frac{\bar{X}-\mu}{\dfrac{\sigma}{\sqrt{n}}} . \frac{\sigma}{S} = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

(2) From the result (c) in Theorem 1.4A,

$$\mathrm{E}\frac{(n-1)S^2}{\sigma^2} = n-1 \quad \Rightarrow \quad \mathrm{E}S^2 = \frac{(n-1)\sigma^2}{n-1} = \sigma^2.$$

and

$$\mathrm{var}\frac{(n-1)S^2}{\sigma^2} = 2n-2 \quad \Rightarrow \quad \mathrm{var}\, S^2 = \frac{\sigma^4}{(n-1)^2}(2n-2) = \frac{2\sigma^4}{(n-1)}. \blacksquare$$

<u>Remarks</u>: (i) $\mathrm{E}S^2 = \sigma^2$ means that $S^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2$ is an <u>unbiased</u> estimator of σ^2 .

(ii) If σ^2 is estimated by $\hat{\sigma}^2 = \frac{1}{n}\sum (X_i - \bar{X})$, then

$$\mathrm{E}\hat{\sigma}^2 = \mathrm{E}\frac{n-1}{n}.\frac{1}{n-1}\sum_i (X_i - \bar{X})^2 = \frac{n-1}{n}\sigma^2 .$$

Thus $\hat{\sigma}^2$ is biased for σ^2 (it underestimates it) and is the reason it is not preferred to S^2.

(iii) $\operatorname{var} S = \mathrm{E}S^2 - \mathrm{E}^2 S = \sigma^2 - \mathrm{E}^2 S > 0 \Rightarrow \mathrm{E}S < \sigma$. Thus although S^2 is unbiased for σ^2, S <u>is actually biased for</u> σ.

1.5 THE EXPONENTIAL FAMILY

Many of the usual parametric families belong to the exponential family (or class). If a one-parameter family of densities $f(x;\theta)$ can be expressed as

$$f(x;\theta) = a(\theta)b(x)\exp\{c(\theta)d(x)\} \tag{1-13}$$

for $-\infty < x < \infty$, for all $\theta \in \Theta$ and suitable functions *a*, *b*, *c*, and *d*, then $f(x;\theta)$ is said to belong to the exponential family (or class).

Example 1.5A

Let $X \sim \text{Poisson}(\lambda)$. Verify if this distribution belongs to the exponential family.

Solution

The p.m.f. of *X* is

$$\begin{aligned} p_X(x;\lambda) &= \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for } x = 0, 1, \ldots \\ &= \frac{e^{-\lambda}\lambda^x}{x!} I\left(x \in \{0, 1, \ldots\}\right) \\ &= e^{-\lambda}\left(\frac{I\left(x \in \{0, 1, \ldots\}\right)}{x!}\right)\exp\ln\lambda^x \\ &= e^{-\lambda}\left(\frac{I\left(x \in \{0, 1, \ldots\}\right)}{x!}\right)\exp\left(x\ln\lambda\right) \end{aligned}$$

Comparing the above with Eq. (1-13), we have $a(\lambda) = e^{-\lambda}$, $b(x) = I\left(x \in \{0, 1, \ldots\}\right)/x!$, $c(\lambda) = \ln\lambda$ and $d(x) = x$.

Example 1.5B

Let $X \sim \text{uniform}(0,\theta)$ where $\theta > 0$. Verify if this distribution belongs to the exponential family.

Solution

We have

$$
\begin{aligned}
f_X(x;\theta) &= \frac{1}{\theta} \quad \text{for} \quad 0 < x < \theta \\
&= \frac{1}{\theta} I(0 < x < \theta)
\end{aligned}
$$

From the above, we see that $I(0 < x < \theta)$ cannot be factored into functions of *x* and θ alone, and $f(x;\theta)$ cannot be written in the exponential form (1-13). Hence, the $\text{uniform}(0,\theta)$ distribution does not belong to the exponential family.

An extension of (1-13) can be made to cover the *k*-parameter exponential family: if a *k*-parameter family of densities $f(x;\theta_1,...,\theta_k)$ can be expressed as

$$
f(x;\theta_1,...,\theta_k) = a(\theta_1,...,\theta_k)b(x)\exp\left\{\sum_{j=1}^{k} c_j(\theta_1,...,\theta_k)d_j(x)\right\} \tag{1-14}
$$

for $-\infty < x < \infty$, for all $\theta \in \Theta$ lying in a generalized rectangle and suitable functions *a*, *b*, c_j, and d_j, then $f(x;\theta_1,...,\theta_k)$ is said to belong to the *k*-parameter exponential family (or class).

Example 1.5C

Let $X \sim N(\mu,\sigma^2)$. Verify if this distribution belongs to the 2-parameter exponential family.

Solution

We have

$$
\begin{aligned}
f_X(x;\mu,\sigma) &= \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \\
&= \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2-2\mu x+\mu^2}{2\sigma^2}\right) \\
&= \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{\mu^2}{2\sigma^2}\right)\exp\left(-\frac{1}{2\sigma^2}x^2+\frac{\mu}{\sigma^2}x\right)
\end{aligned}
$$

Comparing to Eq. (1-14), we have $a(\mu,\sigma)=1/(\sigma\sqrt{2\pi})e^{-\mu^2/(2\sigma^2)}$, $b(x)=1$, $c_1(\mu,\sigma)=-1/(2\sigma^2)$, $d_1(x)=x^2$, $c_2(\mu,\sigma)=\mu/\sigma^2$, and $d_2(x)=x$. Hence, the $N(\mu,\sigma^2)$ distribution belongs to the 2-parameter exponential family.

1.6 THREE IMPORTANT THEOREMS

(a) Lévy's Continuity theorem: Let $X_1, X_2, \ldots$ be a sequence of r.v.'s with X_n having m.g.f. $M_{X_n}(t)$ and let X be a r.v. with m.g.f. $M_X(t)$. Then:

(i) If $\lim_{n\to\infty} M_{X_n}(t) = M_X(t)$ for all $t \in (-a,a)$, then $X_n \xrightarrow{D} X$

(ii) Conversely, if $X_n \xrightarrow{D} X$, then $\lim_{n\to\infty} M_{X_n}(t) = M_X(t)$.

(b) Slutsky's theorem: If $X_n \xrightarrow{D} X$, $A_n \xrightarrow{P} A$, $B_n \xrightarrow{P} B$, where A and B are constants. Then

$A_n X_n + B_n \xrightarrow{D} AX + B$.

Example 1.6A

Consider the Lindeberg-Lévy CLT which states that, under certain conditions,

$$\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \xrightarrow{D} Z \sim N(0,\ 1).$$

Suppose σ is unknown and is replaced by S is Eq. (1-7). What is the asymptotic distribution of $\sqrt{n}(\bar{X}_n - \mu)/S$?

Solution We have

$$\frac{\bar{X}_n - \mu}{S / \sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \cdot \frac{\sigma}{S}$$

Now $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{D} Z \sim N(0,\ 1)$ and $S/\sigma \xrightarrow{P} 1$. Using Slutsky's theorem, $\dfrac{\bar{X}_n - \mu}{S / \sqrt{n}} \xrightarrow{D} Z \sim N(0,\ 1)$.

(c) <u>Cramér's delta theorem:</u> Suppose if $\hat{\theta}_n$ is asymptotically $N(\theta, \sigma^2/n)$, i.e.

$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{D} N(0, \sigma^2)$. Then $f(\hat{\theta}_n)$ can be written as a Taylor series expansion about θ as

$$\begin{aligned} f\left(\hat{\theta}_n\right) &= f(\theta) + \left(\hat{\theta}_n - \theta\right) f^{/}(\theta) + \frac{\left(\hat{\theta}_n - \theta\right)^2}{2!} f^{//}(\theta) + \ldots \\ &= f(\theta) + \left(\hat{\theta}_n - \theta\right) f^{/}(\theta) + o\left(\hat{\theta}_n - \theta\right) \end{aligned} \tag{1-15}$$

where $o(\hat{\theta}_n - \theta)$ is such that $\frac{o(\hat{\theta}_n - \theta)}{\hat{\theta}_n - \theta} \to 0$ as $\hat{\theta}_n \to \theta$. Therefore,

$$\mathrm{E} f\left(\hat{\theta}_n\right) = \mathrm{E} f(\theta) + f^{/}(\theta)\,\mathrm{E}\left(\hat{\theta}_n - \theta\right) + \ldots \approx f(\theta)$$

$$\operatorname{var} f\left(\hat{\theta}_n\right) = 0 + \left\{f^{/}(\theta)\right\}^2 \operatorname{var}\left(\hat{\theta}_n - \theta\right) + \ldots \approx \left\{f^{/}(\theta)\right\}^2 \frac{\sigma^2}{n},$$

assuming $f^{/}(\theta) \neq 0$. Applying Slutsky's theorem to Eq. (1-15) also shows that $f(\hat{\theta}_n)$ is asymptotically normal. Hence, we obtain Cramér's delta theorem:

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{D} N(0, \sigma^2) \;\Rightarrow\; \sqrt{n}\left\{f\left(\hat{\theta}_n\right) - f(\theta)\right\} \xrightarrow{D} N\left[0, \left\{f^{/}(\theta)\right\}^2 \sigma^2\right] \tag{1-16}$$

That is,

$$\hat{\theta}_n \dot{\sim} N\left(\theta, \frac{\sigma^2}{n}\right) \quad \Rightarrow \quad f\left(\hat{\theta}_n\right) \dot{\sim} N\left(f(\theta), \left\{ f'(\theta) \right\}^2 \frac{\sigma^2}{n} \right)$$

In the above, we have assumed that $f'\left(\theta\right) \neq 0$ for any particular θ.

Example 1.6B

Given that $\bar{X}_n$ is asymptotically $N(\mu, \sigma^2 / n)$, find the asymptotic distribution of $1/\bar{X}_n$.

Solution

We have $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} Z \sim N(0, \sigma^2)$ and $f(\bar{X}_n) = 1/\bar{X}_n$. Therefore,

$$f(\bar{X}_n) = f\left(\mu\right) + \left(\bar{X}_n - \mu\right) f'\left(\mu\right) + o\left(\bar{X}_n - \mu\right)$$

and

$$\mathrm{E}\frac{1}{\bar{X}_n} \approx f\left(\mu\right) = \frac{1}{\mu},$$

$$\operatorname{var}\frac{1}{\bar{X}_n} \approx \left\{ f'\left(\mu\right) \right\}^2 \operatorname{var}\bar{X}_n = \left(-\frac{1}{\mu^2}\right)^2 \frac{\sigma^2}{n} = \frac{\sigma^2}{n\mu^4}$$

Therefore, $1/\bar{X}_n$ is asymptotically $N[1/\mu, \sigma^2/(n\mu^4)]$, i.e.

$$\sqrt{n}\left(\frac{1}{\bar{X}_n} - \frac{1}{\mu}\right) \xrightarrow{D} N\left(0, \frac{\sigma^2}{\mu^4}\right)$$

A more general version of Cramér's delta theorem can be proved similarly: if $X_n = a_n(\hat{\theta}_n - \theta) \xrightarrow{D} X$, where a_n is a sequence of constants such that $a_n \to \infty$, then

$$a_n\left\{f\left(\hat{\theta}_n\right)-f\left(\theta\right)\right\}\xrightarrow{D} f'\left(\theta\right)X$$

2 ESTIMATION

2.1 STATISTICAL MODELS

Statistical inference starts by specifying the underlying statistical model, which consists of:

- A random vector $\mathbf{X}=(X_1,...,X_n)\in\chi$ which is observed;
- An unknown parameter vector $\boldsymbol{\theta}=(\theta_1,...,\theta_k)\in\Theta$;
- A function $f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})$ (or $p_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})$) which represents the p.d.f. (or p.m.f.) of $\mathbf{X}$ for each $\boldsymbol{\theta}$.

χ is called the support (or sample space) and Θ is called the parameter space. Note that $\mathbf{X}$ is a sample measure and as such is a r.v. whereas $\boldsymbol{\theta}$ is a population measure and as such is a constant.

Any function $T=T(X_1,...,X_n)$ is called a statistic (note that T is also a r.v.). Note that T must not involve any unknown parameter. When used in the context of providing a numerical value for a parameter, a statistic is called an estimator.

One of the major aims of statistical inference is to use the observed values of suitable T to make conclusions about the unknown $\boldsymbol{\theta}$.

Example 2.1A

Consider the following statistical model: suppose an observation is made on each of $X_1,...,X_{10}$, where each $X_i\overset{iid}{\sim}N(\mu,\sigma^2)$, μ and σ being unknown parameters. Then $\mathbf{X}=(X_1,...,X_{10})$, $\boldsymbol{\theta}=(\mu,\sigma)$, and