Example. $X_1, \ldots, X_n$ i.i.d Poisson(λ), $\lambda > 0$. Consider $T(\mathbf{X}) = \bar{X}$ for estimating λ. Since $E(T) = E(\bar{X}) = \lambda$, T is an unbiased estimator. We also have $Var(T) = Var(\bar{X}) = \frac{Var(X)}{n} = \frac{\lambda}{n}$. Further,

$$\begin{aligned}
f(x|\lambda) &= \exp(-\lambda)\frac{\lambda^x}{x!},\ x = 0, 1, 2, \ldots \\
\log f(x|\lambda) &= -\lambda + x\log(\lambda) - \log(x!), \\
\frac{\partial}{\partial\lambda}\log f(x|\lambda) &= -1 + \frac{x}{\lambda}, \\
\frac{\partial^2}{\partial\lambda^2}\log f(x|\lambda) &= -\frac{x}{\lambda^2},
\end{aligned}$$

so that

$$\begin{aligned}
I_1(\lambda) &= E_\lambda\left[\frac{\partial}{\partial\lambda}\log f(X|\lambda)\right]^2 = Var_\lambda\left[\frac{\partial}{\partial\lambda}\log f(X|\lambda)\right] \\
&= -E_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\log f(X|\lambda)\right] \\
&= \frac{1}{\lambda^2}E_\lambda(X) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}.
\end{aligned}$$

Therefore, $I(\lambda) = I_n(\lambda) = n/\lambda$. This yields the Information bound of

$$Var_\lambda(T) \geq \frac{1}{I(\lambda)} = \frac{\lambda}{n},$$

for any unbiased estimator T. Note that $T(\mathbf{X}) = \bar{X}$ achieves this bound, hence it is UMVUE.

Example. $X \sim$ Poisson(θ), $\theta > 0$. $q(\theta) = \exp(-\theta)$. $I(\theta) = \frac{1}{\theta}$. Consider

$$T(X) = \begin{cases} 1 & \text{if } X = 0; \\ 0 & \text{otherwise.} \end{cases}$$

Then $E(T) = P_\theta(X = 0) = \exp(-\theta) = q(\theta)$ and $Var_\theta(T) = \exp(-\theta)(1 - \exp(-\theta))$ since $T \sim$ Bernoulli($q(\theta)$). The C-R bound on all unbiased estimators U of $q(\theta)$ is

$$Var_\theta(U) \geq \frac{\left(\frac{d}{d\theta}q(\theta)\right)^2}{I(\theta)} = \frac{\exp(-2\theta)}{1/\theta} = \theta\exp(-2\theta) = \text{ C-R }(\theta).$$

Therefore,

$$\begin{aligned}
\frac{Var_\theta(T)}{\text{C-R }(\theta)} &= \frac{\exp(-\theta)(1-\exp(-\theta))}{\theta\exp(-2\theta)} = \frac{1-\exp(-\theta)}{\theta\exp(-\theta)} = \frac{\exp(\theta)-1}{\theta} \\
&= \frac{1+\theta+\theta^2/2+\cdots-1}{\theta} > 1.
\end{aligned}$$

However, X is complete sufficient, hence $T(X)$ is UMVUE of $q(\theta)$.

Confidence Statements

It is not enough to give just an estimate of the parameter of interest, however good the procedure of estimation is. Usually one also wants to know what the likely error of estimation is.

Suppose θ is the parameter of interest, and we have available, a random sample, $X_1, \dots, X_n$ from P_θ. Suppose, further, $\hat{\theta}(X_1, \dots, X_n)$ is an estimator of θ. It is desirable to have an estimate of the magnitude of $\hat{\theta} - \theta$. Typical estimates are asymptotically unbiased. Therefore, an estimate of s.d.$(\hat{\theta})$, called the standard error, s.e.$(\hat{\theta})$, is an indicator of the likely error of estimation of θ by $\hat{\theta}$. This means, $\hat{\theta} \pm$ s.e.$(\hat{\theta})$ is our interval estimate of θ, in the sense that $\hat{\theta}$ is the point estimate but it may be off by s.e.$(\hat{\theta})$.

Example. $\bar{X} \pm \frac{\sigma}{\sqrt{n}}$ or $\bar{X} \pm \frac{s}{\sqrt{n}}$ for μ of $N(\mu, \sigma^2)$; $\hat{p} \pm \sqrt{\hat{p}(1-\hat{p})/n}$ for p of Binomial(n, p); $\hat{\lambda} \pm \sqrt{\hat{\lambda}}$ for λ of Poisson(λ), and so on.

Another more formal approach is through confidence statements.

Interval Estimation

Let $X \sim P_\theta$ and a confidence interval is of interest for $q(\theta)$.

Definition. An interval $[\underline{T}(X), \bar{T}(X)]$, where $\underline{T} \leq \bar{T}$ is a $100(1-\alpha)\%$ confidence interval for $q(\theta)$ if

$$\inf_\theta P_\theta \left\{ \underline{T}(X) \leq \theta \leq \bar{T}(X) \right\} \geq 1 - \alpha.$$

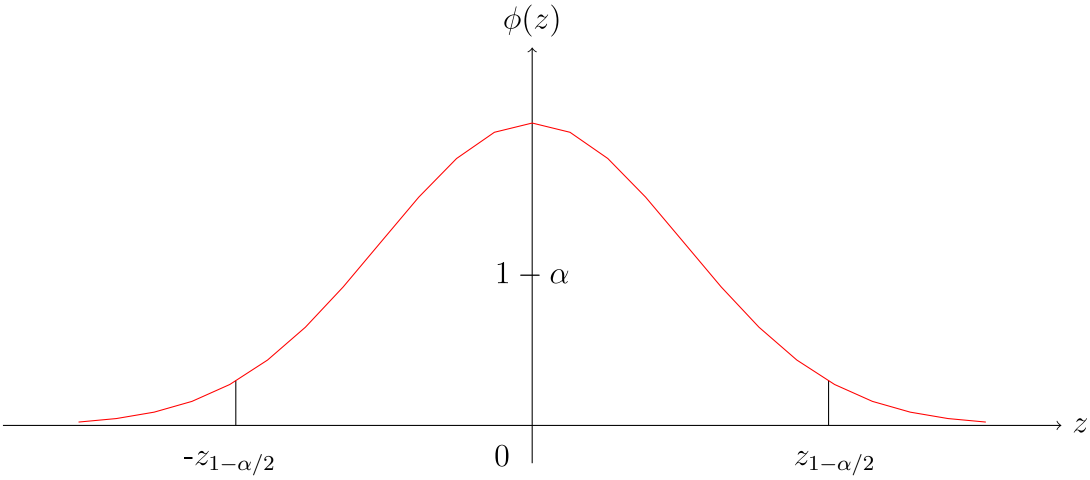
i.e., $P_\theta \left\{ \underline{T}(X) \leq \theta \leq \bar{T}(X) \right\} \geq 1 - \alpha$ for all θ.

Example. Let $X_1, \dots, X_n$ be i.i.d $N(\theta, \sigma^2)$, σ^2 known. Want $[\underline{T}(X), \bar{T}(X)]$ such that $P_\theta \left\{ \underline{T}(X) \leq \theta \leq \bar{T}(X) \right\} \geq 1 - \alpha$ for all θ. Note,

$$Z = \frac{\bar{X} - \theta}{\sigma/\sqrt{n}} \sim N(0, 1) \text{ for all } \theta.$$

Therefore,

$$P_\theta \left(\left| \frac{\bar{X} - \theta}{\sigma/\sqrt{n}} \right| \leq z_{1-\alpha/2} \right) = 1 - \alpha.$$



(For example, if $\alpha = 0.05$, then $z_{1-\alpha/2} = 1.96$.) Thus,

$$\begin{aligned} P_\theta\left(-z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \bar{X} - \theta \leq z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) &= 1-\alpha \text{ for all } \theta, \text{ or} \\ P_\theta\left(\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \theta \leq \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) &= 1-\alpha \text{ for all } \theta. \end{aligned}$$

Therefore, $\left[\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$ is a $100(1-\alpha)\%$ confidence interval for θ.

Example. Let $X_1, \ldots, X_n$ be i.i.d $N(\mu, \sigma^2)$, σ^2 unknown. Now $\theta = (\mu, \sigma^2)$. Want $[\underline{T}(X), \bar{T}(X)]$ such that $P_\theta\left\{\underline{T}(X) \leq \mu \leq \bar{T}(X)\right\} \geq 1-\alpha$ for all $\theta = (\mu, \sigma^2)$. Note, since

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1} \text{ for all } \mu, \sigma^2,$$

$$P_{\mu,\sigma^2}\left(\bar{X} - t_{n-1}(1-\alpha/2)\frac{s}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1}(1-\alpha/2)\frac{s}{\sqrt{n}}\right) = 1-\alpha$$

for all μ, σ^2. Thus, $\left[\bar{X} - t_{n-1}(1-\alpha/2)\frac{s}{\sqrt{n}}, \bar{X} + t_{n-1}(1-\alpha/2)\frac{s}{\sqrt{n}}\right]$ is a $100(1-\alpha)\%$ confidence interval for μ.

Interpretation of confidence statements.

Consider again the confidence interval for μ in $N(\mu, \sigma^2)$ with σ^2 known, which is $\bar{X} \pm z_{1-\alpha/2}\, \sigma/\sqrt{n}$. This means

$$\begin{aligned} & P_{\mu,\sigma^2}\{\mu \in \text{ confidence interval }\} \\ = \;& P_{\mu,\sigma^2}\left\{\bar{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1-\alpha. \end{aligned}$$

In this statement, as in all other areas of classical statistics, μ is a constant, and the probability statement is about $\bar{X}$. So $(1-\alpha)$ is the proportion of times the interval $[\underline{T}, \bar{T}]$ covers μ over repetitions of the experiment and data sets. Let $\alpha = 0.05$. Then, if the interval $\bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$ is used for a large number of data sets in repetitions of the experiment, then in about 95% of the time μ will be inside the interval, and will lie outside rest of the time. If one has a data set with $\bar{X} = 3$, and asks for the probability that μ lies in $3 \pm 1.96\sigma/\sqrt{n}$, the answer isn't 95% but trivially zero or one depending on the value of μ. Though the idea of such intervals is quite old, it was Neyman who formalized them.

The simplest way to generate confidence intervals is to find what Fisher called a pivotal quantity, namely, a real valued function $T(\mathbf{X}, \theta)$ of both $\mathbf{X}$ and θ such that the distribution of $T(\mathbf{X}, \theta)$ does not depend on θ. Suppose then we choose two numbers t_1 and t_2 such that $P_\theta\{t_1 \le T(\mathbf{X}, \theta) \le t_2\} = 1-\alpha$. If for each $\mathbf{x}$, $T(\mathbf{x}, \theta)$ is monotone in θ, say, an increasing function of θ, then we can find $\underline{T}(\mathbf{x})$ and $\bar{T}(\mathbf{x})$ such that $T(\mathbf{x}, \bar{T}(\mathbf{x})) = t_2$ and $T(\mathbf{x}, \underline{T}(\mathbf{x})) = t_1$. Clearly $(\underline{T} \le \theta \le \bar{T})$ iff $t_1 \le T \le t_2$ and hence $\underline{T} \le \theta \le \bar{T}$ with probability $1-\alpha$.

In the normal example, $T(\mathbf{X}, \mu) = \bar{X} - \mu$, the distribution of which is $N(0, \sigma^2/n)$, free of μ.

Example. Let $X_1, \ldots, X_n$ be i.i.d $N(\mu, \sigma^2)$, both μ and σ^2 unknown. How do we construct a confidence interval for σ^2? We need a pivot involving σ^2 only. Consider the MLE of σ^2: $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$, and note that $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2 \sim \sigma^2\chi^2_{n-1}$. Therefore, $S^2/\sigma^2 \sim \chi^2_{n-1}$ can be used as a pivotal statistic. i.e., we can find $c_1 < c_2$ such that

$$\begin{aligned} P_{\sigma^2}\left(c_1 \le \frac{S^2}{\sigma^2} \le c_2\right) &= 1-\alpha \text{ for all } \sigma^2, \\ P_{\sigma^2}\left(\frac{1}{c_2} \le \frac{\sigma^2}{S^2} \le \frac{1}{c_1}\right) &= 1-\alpha \text{ for all } \sigma^2, \\ P_{\sigma^2}\left(\frac{S^2}{c_2} \le \sigma^2 \le \frac{S^2}{c_1}\right) &= 1-\alpha \text{ for all } \sigma^2. \end{aligned}$$

Many choices exist for (c_1, c_2). One may take them to satisfy

$$\frac{\alpha}{2} = P\left(\chi^2_{n-1} \le c_1\right) = 1 - P\left(\chi^2_{n-1} > c_2\right),$$

or take

$$f_{n-1}(c_1) = f_{n-1}(c_2), \text{ and } \int_{c_1}^{c_2} f_{n-1}(x)\,dx = 1 - \alpha.$$

Example. Let $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ be i.i.d. $N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$. Construct confidence interval for $\mu_1 - \mu_2$. Since $\widehat{\mu_1 - \mu_2} = \bar{X} - \bar{Y}$, look for a pivot involving this. Let $D_i = X_i - Y_i$, $i = 1, 2, \ldots, n$. Then D_i are i.i.d $N(\mu_1 - \mu_2, 2\sigma^2(1-\rho) = \sigma^2_D)$. Now, $\bar{D} \sim N(\mu_1 - \mu_2, \sigma^2_D/n)$ independent of $S^2_D = \sum_{i=1}^n (D_i - \bar{D})^2 \sim \sigma^2_D \chi^2_{n-1}$. Therefore,

$$T = \frac{\sqrt{n}\left(\bar{D} - (\mu_1 - \mu_2)\right)}{\sqrt{\sum_{i=1}^n (D_i - \bar{D})^2/(n-1)}} \sim t_{n-1}.$$

Hence, from the previous discussion, $\bar{D} \pm t_{n-1}(1-\alpha/2) s_D/\sqrt{n}$ is a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$. (Here $s^2_D = S^2_D/(n-1)$.)

Example. Let $X_1, \ldots, X_n$ be i.i.d Bernoulli(θ). How do we find $\left[\underline{T}(\mathbf{X}), \bar{T}(\mathbf{X})\right]$ such that

$$P_\theta\left[\underline{T}(\mathbf{X}) \le \theta \le \bar{T}(\mathbf{X})\right] = 1 - \alpha \;\forall\; \theta?$$

$S_n = \sum_{i=1}^n X_i$ is sufficient for θ and $S_n \sim$ Binomial(θ). An exact pivot involving S_n is not available, so it is difficult to construct an exact confidence interval using the above approach. An approximate large sample interval is constructed as follows. If n is large then $\hat{\theta} = S_n/n \sim N(\theta, \theta(1-\theta))$ approximately. Therefore, for large n, approximately,

$$\frac{\hat{\theta} - \theta}{\sqrt{\theta(1-\theta)/n}} \sim N(0,1).$$

In fact, we have, for large n, approximately,

$$\frac{\hat{\theta} - \theta}{\sqrt{\hat{\theta}(1-\hat{\theta})/n}} \sim N(0,1).$$

This gives the usual, large sample, approximate confidence interval:

$$\hat{\theta} \pm z_{1-\alpha/2}\sqrt{\hat{\theta}(1-\hat{\theta})/n}.$$