Module 6

Interval estimation

Topics: Gamma distribution, Confidence interval, Interval estimation of $N(\mu, \sigma^2)$, Interval estimation of other quantities, Introduction to hypothesis testing

Gamma distribution

We begin this module by introducing some new probability distributions that we encounter in statistics. We studied about binomial, geometric, Poisson, exponential, uniform, and normal distributions in random variable. Hyper-geometric distribution was mentioned in convolution. Here, we learn about gamma and associated distributions.

For $r > 0$, gamma function of r is defined as: $\Gamma(r) = \int_0^\infty e^{-u}u^{r-1}du$. Gamma function is a generalization of the factorial function, as explained below. For $r > 1$,

$$\Gamma(r) = \int_0^\infty e^{-u}u^{r-1}du = \left[\frac{e^{-u}}{-1}u^{r-1}\right]_0^\infty - \int_0^\infty \frac{e^{-u}}{-1}(r-1)u^{r-2}du$$
$$= \left[-\lim_{u\to\infty}\frac{u^{r-1}}{e^u} + \frac{0^{r-1}}{e^0}\right] + (r-1)\int_0^\infty e^{-u}u^{r-2}du = (r-1)\Gamma(r-1).$$

For $r = 1$, $\Gamma(1) = \int_0^\infty e^{-u}u^0du = [e^{-u}/-1]_0^\infty = 1$, which is same as 0!. For integer $r \geq 2$, $\Gamma(r) = (r-1)\cdot\Gamma(r-1) = (r-1)(r-2)\cdot\Gamma(r-2) = \cdots = (r-1)(r-2)\cdots 1\cdot\Gamma(1) = (r-1)!$. Therefore, $n! = \Gamma(n+1)$ for $n = 0,1,2,\ldots$.

Using the gamma function, density function of the gamma distribution with positive-valued parameters λ and r, known as rate and shape parameters, is defined as:

$$f(x) = \lambda e^{-\lambda x}\frac{(\lambda x)^{r-1}}{\Gamma(r)} \text{ for } x > 0, \text{and zero otherwise.}$$

$f(x) \geq 0\ \forall x \in \mathbb{R}$ and $\int_{-\infty}^\infty f(x)dx = \int_0^\infty e^{-\lambda x}(\lambda x)^{r-1}\lambda dx/\Gamma(r) = \int_0^\infty e^{-u}u^{r-1}du/\Gamma(r)$, by replacing $\lambda x = u$; $\Rightarrow \int_{-\infty}^\infty f(x)dx = \Gamma(r)/\Gamma(r) = 1$. Therefore, the above function is a valid density. A special case of gamma distribution is the Erlang distribution, which is encountered when r is integer valued. If $r = 1$, then the gamma/Erlang distribution becomes exponential distribution. Sum of *iid* exponential random variables has Erlang distribution.

Let us determine moment generating function of $X \sim Gamma(\lambda, r)$.

$$m_X(s) = E[e^{sX}] = \int_0^\infty e^{sx}\lambda e^{-\lambda x}\frac{(\lambda x)^{r-1}}{\Gamma(r)}dx = \frac{\lambda^r}{\Gamma(r)}\int_0^\infty e^{-(\lambda-s)x}x^{r-1}dx$$
$$= \frac{\lambda^r}{\Gamma(r)}\int_0^\infty e^{-u}\left(\frac{u}{\lambda-s}\right)^{r-1}\frac{du}{\lambda-s}, \text{by replacing } (\lambda-s)x = u$$
$$= \frac{\lambda^r}{\Gamma(r)(\lambda-s)^r}\int_0^\infty e^{-u}u^{r-1}du = \frac{\lambda^r\Gamma(r)}{\Gamma(r)(\lambda-s)^r}, \text{provided } u > 0$$

$$\Rightarrow m_X(s) = \left(\frac{\lambda}{\lambda - s}\right)^r \text{ for } u = (\lambda - s)x > 0 \equiv s < \lambda$$

$m_X'(s) = r\lambda^r/(\lambda - s)^{r+1}$ and $m_X''(s) = r(r+1)\lambda^r/(\lambda - s)^{r+2}$. Then $E[X] = m_X'(0) = r/\lambda$ and $Var(X) = E[X^2] - E^2[X] = m_X''(0) - (r/\lambda)^2 = r(r+1)/\lambda^2 - r^2/\lambda^2 = 1/\lambda$. Observe the resemblance with mean and variance of exponential distribution.

If we add independent gamma random variables with identical rate parameters λ and arbitrary shape parameters $r_1, r_2, \ldots, r_n$, then the sum follows gamma distribution with parameters λ and $r_1 + r_2 + \cdots + r_n$. Let $X_i \sim Gamma(\lambda, r_i)$ for $i = 1,2, \ldots, n$ be independent.

$$\text{Then } m_{X_i}(s) = E[e^{sX_i}] = \left(\frac{\lambda}{\lambda - s}\right)^{r_i} \text{ for } s < \lambda \text{ and } i = 1,2, \ldots, n$$

$$\Rightarrow m_{\sum_{i=1}^n X_i}(s) = E\left[e^{s(X_1+X_2+\cdots+X_n)}\right] = E[e^{sX_1}]E[e^{sX_2}]\cdots E[e^{sX_n}] \text{ due to independence}$$

$$= \left(\frac{\lambda}{\lambda - s}\right)^{r_1+r_2+\cdots+r_n} \quad \text{for } s < \lambda$$

Then by the uniqueness theorem, $X_1 + X_2 + \cdots + X_n \sim Gamma\,(\lambda, r_1 + r_2 + \cdots + r_n)$. Note that this also proves that the sum of n *iid* exponential random variables with parameter λ has Erlang distribution with parameters λ and n.

χ^2 distribution: Another special case of Gamma distribution is the χ^2 distribution, which is encountered when $\lambda = 1/2$ and $r = n/2$ where $n = 1,2,3, \ldots$. So, χ^2 distribution has single parameter n, known as its degree of freedom. Since $\chi^2(n) \equiv \Gamma(1/2\,, n/2)$,

$$f_{\chi^2(n)} = e^{-x/2}\frac{x^{n/2-1}}{2^{n/2}\Gamma(n/2)} \text{ for } x > 0, \text{and zero otherwise.}$$

Also, $E[\chi^2(n)] = n$ and $Var\big(\chi^2(n)\big) = 2n$. The distribution is called χ^2, not χ, because it is the distribution of the sum of squared *iid* standard normal random variables. Let $Z \sim N(0,1)$. Then $Z^2 \sim \chi^2(1)$, as shown below. For $x > 0$,

$$F_{Z^2}(x) = P(Z^2 \le x) = P\big(-\sqrt{x} \le Z \le \sqrt{x}\big) = \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$$\Rightarrow f_{Z^2}(x) = F_{Z^2}'(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{x})^2}{2}} \cdot \frac{1}{2\sqrt{x}} - \frac{1}{\sqrt{2\pi}} e^{-\frac{(-\sqrt{x})^2}{2}} \cdot \frac{-1}{2\sqrt{x}} = \frac{e^{-x/2}}{\sqrt{2\pi x}} \text{ for } x > 0$$

$$\text{Now, } \Gamma(1/2) = \int_0^\infty e^{-u} u^{\frac{1}{2}-1} du = \int_0^\infty e^{-\frac{v^2}{2}} \left(\frac{v^2}{2}\right)^{-1/2} v dv\,, \text{by replacing } u = \frac{v^2}{2}$$

$$= \sqrt{2}\int_0^\infty e^{-\frac{v^2}{2}} dv = 2\sqrt{\pi}\int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} dv = 2\sqrt{\pi} \cdot \frac{1}{2} = \sqrt{\pi}$$

$$\Rightarrow f_{Z^2}(x) = \frac{e^{-x/2}}{\sqrt{2x}\Gamma(1/2)} = e^{-x/2}\frac{x^{1/2-1}}{2^{1/2}\Gamma(n/2)} \text{ for } x > 0 \;\Rightarrow f_Z^2 = f_{\chi^2(1)}, \text{as claimed.}$$

Let $Z_i \sim N(0,1)$ for $i = 1,2,\ldots,n$ be independent. Then $Z_i^2 \sim \chi^2(1) \equiv Gamma(1/2\,,1/2)$ for $i = 1,2,\ldots,n$ are independent. Then $Z_1^2 + Z_2^2 + \cdots = Z_n^2 \sim Gamma(1/2\,,n/2) \equiv \chi^2(n)$, as mentioned earlier. So, the sum of squared *iid* standard normal random variables follow the χ^2 distribution with n degrees of freedom. In statistics, we need quantiles of χ^2 distribution. We cannot get them analytically. There are tables; familiarize yourself with them. The tables provide quantile values for smaller values of n. For larger n, due to the central limit theorem, χ^2 distribution can be approximated by $N(n, 2n)$.

***t* and *F* distributions**: Let $Z \sim N(0,1)$ and $V \sim \chi^2(n)$ be independent. Then $T = Z/\sqrt{V/n}$ is defined to follow t distribution with parameter n (degrees of freedom).

$$F_{\sqrt{V/n}}(v) = P\left(\sqrt{V/n} \le v\right) = P(V \le nv^2) = \int_0^{nv^2} \frac{e^{-x/2}x^{n/2-1}}{2^{n/2}\Gamma(n/2)}dx \text{ for } v > 0$$

$$\Rightarrow f_{\sqrt{V/n}}(v) = F'_{\sqrt{V/n}}(v) = \frac{e^{-nv^2/2}(nv^2)^{n/2-1}}{2^{n/2}\Gamma(n/2)} \cdot 2nv = \frac{e^{-nv^2/2}n^{n/2}v^{n-1}}{2^{n/2-1}\Gamma(n/2)} \text{ for } v > 0$$

Now, using the quotient formula introduced in the end of random vectors,

$$f_T(t) = \int_{-\infty}^{\infty} |v| f_Z(vt) f_{\sqrt{V/n}}(v) dv = \int_0^{\infty} v \cdot \frac{e^{-(vt)^2/2}}{\sqrt{2\pi}} \cdot \frac{e^{-nv^2/2}n^{n/2}v^{n-1}}{2^{n/2-1}\Gamma(n/2)} dv$$

$$= \frac{\sqrt{2/\pi}}{\Gamma(n/2)} \int_0^{\infty} e^{-\frac{(t^2+n)v^2}{2}} \left(\frac{nv^2}{2}\right)^{\frac{n}{2}} dv = \frac{\sqrt{n/\pi}}{\Gamma(n/2)} \int_0^{\infty} e^{-\frac{(t^2+n)v^2}{2}} \left(\frac{nv^2}{2}\right)^{\frac{n-1}{2}} vdv$$

$$= \frac{\sqrt{n/\pi}}{\Gamma(n/2)} \int_0^{\infty} e^{-u} \left(\frac{nu}{t^2+n}\right)^{\frac{n-1}{2}} \frac{du}{t^2+n}, \text{by substituting } \frac{(t^2+n)v^2}{2} = u\ (> 0)$$

$$= \frac{\sqrt{1/n\pi}}{\Gamma(n/2)} \left(\frac{n}{t^2+n}\right)^{\frac{n+1}{2}} \int_0^{\infty} e^{-u} u^{\frac{n+1}{2}-1} du = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \quad \forall t \in \mathbb{R}$$

In statistics, we need quantiles of t distribution We cannot get them analytically. There are tables; familiarize yourself with them. Tables provide quantile values for smaller values of n. For larger n, t distribution can be approximated by $N(0,1)$. This is because, $V \approx N(n, 2n) \equiv V/n \approx N(1, 2/n)$ for sufficiently large n, by the central limit theorem. Since the variance is vanishing with increasing n, in non-technical terms, V/n converges to its mean, i.e., 1, as n increases, implying convergence of $T = Z/\sqrt{V/n}$ to $Z \sim N(0,1)$.

Let $U \sim \chi^2(m)$ and $V \sim \chi^2(n)$ be independent. Then $F = (U/m)/(V/n)$ is said to follow F distribution with parameters m and n (degrees of freedom). It's important to write the order of m and n correctly. Using the quotient formula, one can show that

$$f_{F(m,n)}(x) = \frac{\Gamma(m/2 + n/2)}{\Gamma(m/2)\Gamma(n/2)} \frac{(m/n)^{m/2}x^{m/2-1}}{(1 + mx/n)^{(m+n)/2}} \text{ for } x > 0$$

In statistics, we need quantiles of F distribution We cannot get them analytically. There are tables; familiarize yourself with them. It shall be noted that $(V/n)/(U/m) = 1/F$ follows F distribution with parameters n and m (see the order). Then $F_{F(n,m)}(x) = P(1/F \le x) = P(F \ge 1/x) = 1 - F_{F(m,n)}(1/x)$ for all $x > 0$. In terms of quantiles,

$$\alpha = F_{F(n,m)}\big(q_{\alpha|n,m}\big) = 1 - F_{F(m,n)}\big(1/q_{\alpha|n,m}\big) \Rightarrow F_{F(m,n)}\big(1/q_{\alpha|n,m}\big) = 1 - \alpha$$
$$\Rightarrow 1/q_{\alpha|n,m} = q_{(1-\alpha)|m,n} \Rightarrow q_{\alpha|n,m} = 1/q_{(1-\alpha)|m,n} \ \forall\alpha \in (0,1)$$

Confidence interval

Let $X \sim N(\mu, 100)$ with unknown μ. This situation is somewhat artificial as the variance is known while the mean is unknown. We are considering it to illustrate the idea of confidence interval. Let $X_1, X_2, \dots, X_n$ denote a simple random sample of X. Then $\hat{\mu} = \bar{X} = \sum_{i=1}^{n} X_i/n$ is the 'natural' estimator of μ. It's also the maximum likelihood estimator. Now, consider a realization of the simple random sample to be $36,54,59,61,59$. Then $\hat{\mu} = \bar{X} = 53.8$ is our estimate for μ. Note that we cannot tell $\mu = 53.8$, but we can say that μ, whatever it may be, is 'not too far' from 53.8. The meaning of 'not too far' is captured by bias, efficiency, and consistency. We discussed these aspects of an estimator in the previous module. Here, we measure the idea of 'not too far' more preciously, through confidence interval.

Let us determine the probability that $\hat{\mu} = \bar{X}$ is within ± 2 of μ in the above example. Since $X \sim N(\mu, 100)$ and $X_1, X_2, \dots, X_5$ are independent random variables with the distribution of X, then $\bar{X} = \sum_{i=1}^{5} X_i/5 \sim N(\mu, 100/5 = 20)$ and $(\bar{X} - \mu)/\sqrt{20} \sim N(0,1)$. Therefore,

$$P(\mu - 2 \le \bar{X} \le \mu + 2) = P\left(\frac{-2}{\sqrt{20}} \le \frac{\bar{X} - \mu}{\sqrt{20}} \le \frac{2}{\sqrt{20}}\right) = \Phi\left(\frac{1}{\sqrt{5}}\right) - \Phi\left(\frac{-1}{\sqrt{5}}\right) = 0.345.$$

We can determine $P(\mu - \delta \le \bar{X} \le \mu + \delta)$ for any $\delta > 0$. We can also determine $P(\mu - \delta_l \le \bar{X} \le \mu + \delta_r)$, $P(\mu - \delta_l \le \bar{X})$, and $P(\bar{X} \le \mu + \delta_r)$ for any δ_l, δ_r. *Obtain* $P(\mu - 5 \le \bar{X} \le \mu + 5)$, $P(\mu - 3 \le \bar{X} \le \mu + 4)$, $P(\mu - 1 \le \bar{X})$, $P(\bar{X} \le \mu + 1)$ *for the above example*. So, we can calculate the probability that our estimate is within a range around the quantity of interest. This is a more precise measure of the idea of 'not too far'.

Now, let us ask the converse question, i.e., determine a range that would 'hold' the quantity of interest with certain probability (or confidence). This range is called confidence interval. Let us determine a 90% confidence interval for the example mentioned above. We know that $(\bar{X} - \mu)/\sqrt{20} \sim N(0,1)$. From the normal table, we obtain 5% and 95% quantiles of $N(0,1)$: $z_{0.05} = -1.645$ and $z_{0.95} = 1.645$. Then

$$P\big(z_{0.05} \le (\bar{X} - \mu)/\sqrt{20} \le z_{0.95}\big) = \Phi(z_{0.95}) - \Phi(z_{0.05}) = 0.95 - 0.05 = 0.9$$
$$\Rightarrow P\big(\bar{X} - z_{0.95}\sqrt{20} \le \mu \le \bar{X} - z_{0.05}\sqrt{20}\big) = 0.9 \text{ (or 90\%)}$$

So, $\big[53.8 - 1.645\sqrt{20}, 53.8 + 1.645\sqrt{20}\big] = [46.44, 61.16]$ is a 90% confidence interval for μ. The above calculation also suggests that we could have chosen $z_{0.04}, z_{0.94}$ or $z_{0.06}, z_{0.96}$ or

$z_0, z_{0.9}$ or $z_{0.1}, z_1$ or some other pair (in place of $z_{0.05}, z_{0.95}$) for constructing the confidence interval. *Determine* 90% *confidence interval with these alternate choices*. In general, for a $(1-\alpha)\times 100\%$ confidence interval, where $\alpha \in (0,1)$, we can choose $z_{\alpha/2+\epsilon}, z_{1-\alpha/2+\epsilon}$ for any $\epsilon \in [-\alpha/2, \alpha/2]$. Typically, $\epsilon = 0$, i.e., $z_{\alpha/2}$, $z_{1-\alpha/2}$, is our choice, because it leads to smaller confidence interval. *Verify this for the above example*. Sometimes, we are interested in $z_0, z_{1-\alpha}$ or z_α, z_1 (i.e., $\epsilon = -\alpha/2$ or $\alpha/2$ respectively). These are respectively known as right and left -sided confidence intervals.

Confidence interval is an alternate way of estimating a quantity of interest. Here, we obtain a range that 'holds' the quantity of interest with a specified confidence level. The key is to find a point estimator (MLE or otherwise) whose distribution is known and involves the quantity of interest. In the above example, $\bar{X} \sim N(\mu, 20)$ is that point estimator. Then we use quantiles of the distribution to write a 'probability statement', which eventually leads to the confidence interval. In the above example, the probability statement is: $P\left(z_{0.05} \le (\bar{X}-\mu)/\sqrt{20} \le z_{0.95}\right) = 0.9$, which lead to $\left[\bar{X} - z_{0.95}\sqrt{20}, \bar{X} - z_{0.05}\sqrt{20}\right]$ as the confidence interval.

Let us place ourselves before the data collection, i.e., before realization of the simple random sample, and construct a $(1-\alpha)\times 100\%$ confidence interval for μ of $N(\mu, \sigma^2)$ with known σ^2. We choose $\bar{X} \sim N(\mu, \sigma^2/n)$ to be our point estimator. Then

$$P\left(z_{\alpha/2} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le z_{1-\alpha/2}\right) = \Phi\left(z_{1-\alpha/2}\right) - \Phi\left(z_{\alpha/2}\right) = (1-\alpha/2) - \alpha/2 = 1-\alpha$$

$$\Rightarrow P\left(\bar{X} - z_{1-\alpha/2}\left(\sigma/\sqrt{n}\right) \le \mu \le \bar{X} - z_{\alpha/2}\left(\sigma/\sqrt{n}\right)\right) = 1-\alpha$$

Due to symmetry, $-z_{\alpha/2} = z_{1-\alpha/2}$. Then the confidence interval is $\left[\bar{X} - z_{1-\alpha/2}\left(\sigma/\sqrt{n}\right), \bar{X} + z_{1-\alpha/2}\left(\sigma/\sqrt{n}\right)\right]$. *Show that* $(1-\alpha)\times 100\%$ *right and left-sided confidence interval of* μ *are* $\left(-\infty, \bar{X} + z_{1-\alpha}\left(\sigma/\sqrt{n}\right)\right]$ *and* $\left[\bar{X} - z_{1-\alpha}\left(\sigma/\sqrt{n}\right), \infty\right)$. Observe that constructing a $(1-\alpha)\times 100\%$ confidence interval of θ is essentially constructing functions $l(X_1, X_2, \dots, X_n, \alpha)$ and $r(X_1, X_2, \dots, X_n, \alpha)$ such that $P\left(l(\cdots) \le \theta \le r(\cdots)\right) = 1-\alpha$.

Before we conclude this section, let us note a couple of points. In the above calculations, we took $\Phi(z_\alpha) = \alpha\ \forall\ \alpha \in (0,1)$. This is true because Φ is distribution function of a continuous random variable, which guarantees $F(q_\alpha) = \alpha$ for all α; see the definition of quantile. It is not the case with discrete random variables. We chose $\bar{X}$ as the point estimator for μ in the above calculations. We could have chosen something else, say X_1. Since $X_1 \sim (\mu, \sigma^2)$, the above principle would give us $\left[X_1 - z_{1-\alpha/2}\sigma, X_1 + z_{1-\alpha/2}\sigma\right]$ as the confidence interval for μ. Even though it's a valid one, it is wider than the one we got from $\bar{X}$. Typically, a good point estimator leads to a smaller confidence interval.

Interval estimation of $N(\mu, \sigma^2)$

In the previous section, we found confidence interval for μ of $N(\mu, \sigma^2)$, where σ^2 is known.

This is somewhat artificial as μ is unknown and σ^2 is known. Let us consider a more realistic situation. Let us determine confidence intervals for both μ and σ^2 of $N(\mu, \sigma^2)$. MLE of μ is $\bar{X} \sim N(\mu, \sigma^2/n)$; this is no more useful as σ^2 is unknown. However, if we can think of some quantity whose distribution does not depend on σ^2, then it may be useful. For constructing such a quantity, we would need the following result, which is regarded as the foundation for small sample theory. We will discuss the case of large sample later.

Result: Let $X_1, X_2, \dots, X_n$ denote a simple random sample of $N(\mu, \sigma^2)$. Let $\bar{X} = \sum_{i=1}^n X_i/n$ and $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ denote sample mean and variance. Then

(a) $\bar{X}$ and S^2 are independent random variables,

(b) $(n-1)S^2/\sigma^2$ has $\chi^2(n-1)$ distribution.

$\bar{X}$ and S^2 both are functions of the same set of random variables $X_1, X_2, \dots, X_n$. Therefore, it is rather surprising that these are independent. This happens in rare cases when the functions capture unrelated aspects of the random variables. Consider A_1 and A_2 to be independent and uniformly distributed in $\{1,2,3\}$. Let $B = \max\{A_1, A_2\}$ and $C = 1$ if $A_1 > A_2$, 0 if $A_1 < A_1$, $Ber(0.5)$ if $A_1 = A_2$. Joint and marginal mass functions of B, C are shown below.

$p_{B,C}(b,c)$	$c = 0$	$c = 1$	$p_B(b)$
$b = 1$	1/18	1/18	1/9
$b = 2$	3/18	3/18	3/9
$b = 3$	5/18	5/18	5/9
$p_C(c)$	1/2	1/2	1

Observe that B and C are independent despite both being functions of A_1, A_2. A similar situation arises with $\bar{X}$ and S^2. They respectively capture μ and σ^2 of $N(\mu, \sigma^2)$, and μ, σ^2 of $N(\mu, \sigma^2)$ are unrelated. A formal proof is beyond our scope.

Given the independence of $\bar{X}$ and S^2, we can prove that $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$. Let us define $Z_i = (X_i - \mu)/\sigma$ for $i = 1,2,\dots,n$. Then $Z_1, Z_2, \dots, Z_n$ are *iid* standard normal random variables. Let us represent $(n-1)S^2/\sigma^2$ in terms of $Z_1, Z_2, \dots, Z_n$.

$$\bar{Z} := \frac{\sum_{i=1}^n Z_i}{n} = \frac{\sum_{i=1}^n (X_i - \mu)}{n\sigma} = \frac{\bar{X} - \mu}{\sigma} \Rightarrow \bar{X} = \mu + \bar{Z}\sigma$$

$$\Rightarrow \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu - \bar{Z}\sigma}{\sigma}\right)^2 = \sum_{i=1}^n (Z_i - \bar{Z})^2 = \sum_{i=1}^n Z_i^2 - n\bar{Z}^2$$

So, we have: $\sum_{i=1}^n Z_i^2 = n\bar{Z}^2 + (n-1)S^2/\sigma^2$. Since $Z_1, Z_2, \dots, Z_n$ are independent standard normal random variables, $\sum_{i=1}^n Z_i^2 \sim \chi^2(n) \equiv Gamma(1/2, n/2)$. Then the *mgf* of $\sum_{i=1}^n Z_i^2$: $m_1(s) = (0.5/(0.5-s))^{n/2} = (1-2s)^{-n/2}$ for $s < 0.5$. Now, $\sqrt{n}\bar{Z} = (\bar{X} - \mu)/(\sigma/\sqrt{n}) \sim N(0,1) \Rightarrow n\bar{Z}^2 \sim \chi^2(1)$. Its *mgf*, $m_2(s) = (1-2s)^{-1/2}$ for $s < 0.5$. Let $m_3(s)$ denote the *mgf* of $(n-1)S^2/\sigma^2$. Since $\bar{X} = \mu + \bar{Z}\sigma$ and S^2 are independent, $n\bar{Z}^2$ and $(n-1)S^2/\sigma^2$ are independent too. Then the *mgf* of $n\bar{Z}^2 + (n-1)S^2/\sigma^2$ is: $m_2(s)m_3(s)$, which must be equal to $m_1(s)$ as $\sum_{i=1}^n Z_i^2 = n\bar{Z}^2 + (n-1)S^2/\sigma^2$.

$\Rightarrow (1-2s)^{-1/2} m_3(s) = (1-2s)^{-n/2} \Rightarrow m_3(s) = (1-2s)^{-(n-1)/2}$ for $s < 0.5$

$\Rightarrow (n-1)S^2/\sigma^2 \sim \chi^2(n-1)$ by the uniqueness theorem of mgf.

With the above result, we can construct confidence intervals for both μ and σ^2 of $N(\mu, \sigma^2)$. When σ^2 was known, we considered $(\bar{X}-\mu)/(\sigma/\sqrt{n}) \sim N(0,1)$ for the confidence interval of μ. Since σ^2 is unknown here, let us consider $(\bar{X}-\mu)/(S/\sqrt{n})$.

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{(\bar{X}-\mu)/(\sigma/\sqrt{n})}{S/\sigma} = \frac{(\bar{X}-\mu)/(\sigma/\sqrt{n})}{\sqrt{\{(n-1)S^2/\sigma^2\}/(n-1)}} \sim T(n-1),$$

as the numerator and denominator are independent and $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$. From this, we can obtain confidence interval for μ. Let t_α for $\alpha \in (0,1)$ denote α-th quantile of t distribution with $n-1$ degrees of freedom. Then

$$P\left(t_{\alpha/2} \le \frac{\bar{X}-\mu}{S/\sqrt{n}} \le t_{1-\alpha/2}\right) = \mathrm{F_T}(t_{1-\alpha/2}) - \mathrm{F_T}(t_{\alpha/2}) = 1-\alpha$$
$$\Rightarrow P\left(\bar{X} - t_{1-\alpha/2}(S/\sqrt{n}) \le \mu \le \bar{X} - t_{\alpha/2}(S/\sqrt{n})\right) = 1-\alpha$$

Due to symmetry, $-t_{\alpha/2} = t_{1-\alpha/2}$. Then the $(1-\alpha)\times 100\%$ confidence interval for μ is: $[\bar{X} - t_{1-\alpha/2}(S/\sqrt{n}), \bar{X} + t_{1-\alpha/2}(S/\sqrt{n})]$. Similarly, one can establish that the right and left-sided confidence intervals are: $(-\infty, \bar{X} + t_{1-\alpha}(S/\sqrt{n})]$ and $[\bar{X} - t_{1-\alpha}(S/\sqrt{n}), \infty)$.

We can obtain confidence interval for σ^2 from $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$. Let χ^2_α for $\alpha \in (0,1)$ denote α-th quantile of $\chi^2(n-1)$ distribution. Then

$$P\left(\chi^2_{\alpha/2} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{1-\alpha/2}\right) = \mathrm{F}_{\chi^2}(\chi^2_{1-\alpha/2}) - \mathrm{F}_{\chi^2}(\chi^2_{\alpha/2}) = 1-\alpha$$
$$\Rightarrow P\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{\alpha/2}}\right) = 1-\alpha$$

So, a $(1-\alpha)\times 100\%$ confidence interval for σ^2 is: $[(n-1)S^2/\chi^2_{1-\alpha/2}, (n-1)S^2/\chi^2_{\alpha/2}]$. In a similar manner, one can show that the right and left-sided confidence intervals are: $(-\infty, (n-1)S^2/\chi^2_\alpha]$ and $[(n-1)S^2/\chi^2_{1-\alpha}, \infty)$. It shall be noted that χ^2_α is not square of χ_α, it's the α-th quantile of $\chi^2(n-1)$ distribution.

Sample sizes less than 30 are typically regarded as small sample and the above confidence intervals apply. For large samples, i.e., when $n \ge 30$, we can use normal approximation of t and χ^2 distributions. For large n, $T(n-1) \approx N(0,1)$ and $\chi^2(n-1) \approx N(n-1, 2(n-1))$. Then $t_\alpha \approx z_\alpha$ and $\chi^2_\alpha = (n-1) + z_\alpha\sqrt{2(n-1)}$ for all $\alpha \in (0,1)$.

Interval estimation of other quantities

Let us obtain confidence intervals for parameters associated with distributions other than the normal distribution. Let us begin with λ of $Exp(\lambda)$. MLE of λ is: $\hat{\lambda} = 1/\bar{X}$, but it's not easy to obtain distribution of $1/\bar{X}$. Let us consider the quantity $n\bar{X} = \sum_{i=1}^{n} X_i$. It follows $Erl(\lambda, n)$

distribution, as $\sum_{i=1}^{n} X_i$ is sum of n independent $Exp(\lambda)$ random variables. However, quantile of Erlang distribution is not easy to obtain. So, we tweak the quantity further. Let us consider $2\lambda n\bar{X}$. Its moment generating function tells us that $2\lambda n\bar{X} \sim \chi^2(2n)$.

$$m_{2\lambda n\bar{X}}(s) = E\left[e^{s\cdot 2\lambda \sum_{i=1}^{n} X_i}\right] = \left[e^{t\sum_{i=1}^{n} X_i}\right] = m_{\sum_{i=1}^{n} X_i}(t), \text{where } t = 2\lambda s$$

$$\text{Since } \sum_{i=1}^{n} X_i \sim Erl(\lambda, n) \equiv Gamma(\lambda, n), m_{\sum_{i=1}^{n} X_i}(t) = \left(\frac{\lambda}{\lambda - t}\right)^n \text{ for } t < \lambda$$

$$\Rightarrow m_{2\lambda n\bar{X}}(s) = \left(\frac{\lambda}{\lambda - t}\right)^n = \left(\frac{\lambda}{\lambda - 2\lambda s}\right)^n = (1-2s)^{-n} \text{ for } 2\lambda s < \lambda \equiv s < 1/2$$

Since $m_{\chi^2(2n)}(s) = (1-2s)^{-n}$ for $s < 1/2$, we conclude that $2\lambda n\bar{X} \sim \chi^2(2n)$. Using this distribution, we can obtain $(1-\alpha) \times 100\%$ confidence interval for λ of $Exp(\lambda)$. Let χ^2_α for $\alpha \in (0,1)$ denote α-th quantile of $\chi^2(2n)$ distribution. Then

$$P\left(\chi^2_{\alpha/2} \leq 2\lambda n\bar{X} \leq \chi^2_{1-\alpha/2}\right) = \mathrm{F}_{\chi^2}\left(\chi^2_{1-\alpha/2}\right) - \mathrm{F}_{\chi^2}\left(\chi^2_{\alpha/2}\right) = 1 - \alpha$$

$$\Rightarrow P\left(\frac{\chi^2_{\alpha/2}}{2n\bar{X}} \leq \lambda \leq \frac{\chi^2_{1-\alpha/2}}{2n\bar{X}}\right) = 1 - \alpha$$

So, a $(1-\alpha) \times 100\%$ confidence interval for λ is: $\left[\chi^2_{\alpha/2}/2n\bar{X}, \chi^2_{1-\alpha/2}/2n\bar{X}\right]$. Similarly, one can show that the right and left-sided confidence intervals of λ are: $(-\infty, \chi^2_{1-\alpha}/2n\bar{X}]$ and $[\chi^2_\alpha/2n\bar{X}, \infty)$. For large n, $\chi^2(2n) \approx N(2n, 4n)$. Then $\chi^2_\alpha = 2n + z_\alpha\sqrt{4n}$ for all α.

Let us consider μ of $Pois(\mu)$. MLE of μ is: $\hat{\mu} = \bar{X}$, but it's not easy to obtain distribution of $\bar{X}$. Let us consider the quantity $n\bar{X} = \sum_{i=1}^{n} X_i$. It follows $Pois(n\mu)$ distribution, as the sum of independent Poisson random variables follows Poisson distribution with parameters added together. However, quantile of Poisson distribution is not easy to obtain. Here, we rely on the approximation due to the central limit theorem. For sufficiently large n, $\sum_{i=1}^{n} X_i \approx N(n\mu, n\mu)$ $\equiv \bar{X} \approx N(\mu, \mu/n) \equiv (\bar{X} - \mu)/\sqrt{\mu/n} \sim N(0,1)$. Then

$$P\left(z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sqrt{\mu/n}} \leq z_{1-\alpha/2}\right) \approx \Phi\left(z_{1-\alpha/2}\right) - \Phi\left(z_{\alpha/2}\right) = 1 - \alpha$$

$$\Rightarrow P\left(-z_{1-\alpha/2} \leq \frac{\bar{X} - \mu}{\sqrt{\mu/n}} \leq z_{1-\alpha/2}\right) = P\left(\frac{(\bar{X} - \mu)^2}{\mu/n} \leq z^2_{1-\alpha/2}\right) \approx 1 - \alpha$$

$$\Rightarrow P\left(\mu^2 - \left(2\bar{X} + \frac{z^2}{n}\right)\mu + \bar{X}^2 \leq 0\right) \approx 1 - \alpha, \text{where } z = z_{1-\alpha/2}$$

Solutions of the quadratic equation $\mu^2 - (2\bar{X} + z^2/n)\mu + \bar{X}^2 = 0$ are:

$$\mu = \frac{(2\bar{X} + z^2/n) \pm \sqrt{(2\bar{X} + z^2/n)^2 - 4\bar{X}^2}}{2} = \left(\bar{X} + \frac{z^2}{2n}\right) \pm \frac{z}{\sqrt{n}}\sqrt{\bar{X} + \frac{z^2}{4n}} \approx \bar{X} \pm z\sqrt{\frac{\bar{X}}{n}},$$

as $z^2/2n$ or $z^2/4n$ are insignificant compared to $\bar{X}$ for large n. Note that both the solutions are real. From the theory of quadratic equations, if μ is within these two solutions, then $\mu^2 - (2\bar{X} + z^2/n)\mu + \bar{X}^2 \le 0$ and vice-versa. Then

$$P\left(\bar{X} - z_{1-\alpha/2}\sqrt{\bar{X}/n} \le \mu \le \bar{X} + z_{1-\alpha/2}\sqrt{\bar{X}/n}\right) \approx 1-\alpha$$

Thus, for large n, $\left[\bar{X} - z_{1-\alpha/2}\sqrt{\bar{X}/n}\,, \bar{X} + z_{1-\alpha/2}\sqrt{\bar{X}/n}\right]$ is a $(1-\alpha) \times 100\%$ 'approximate' confidence interval for μ of $Pois(\mu)$.

Let us consider p of $Ber(p)$. MLE of p is: $\hat{p} = \bar{X}$, but it's not easy to obtain distribution of $\bar{X}$. We can consider the quantity $n\bar{X} = \sum_{i=1}^{n} X_i$, which follows $Bin(p,n)$ distribution. However, quantile of binomial distribution is not easy to obtain. For large n, $\sum_{i=1}^{n} X_i \approx N\big(np, np(1-p)\big)$ by the central limit theorem. With this, we can proceed like the case of $Pois(\mu)$. Let the use a different approximation this time. Let us consider asymptotic normality of MLEs. Then for large n, $\hat{p} = \bar{X} \approx N(p, 1/n\mathbb{I}(\mathrm{p}))$, where

$$\mathbb{I}(\theta) = E\left[\left(\frac{f'(X|p)}{f(X|p)}\right)^2\right] = E\left[\left(\frac{Xp^{X-1}(1-p)^{1-X} - (1-X)p^X(1-p)^{-X}}{p^X(1-p)^{1-X}}\right)^2\right]$$
$$= E\left[\left(\frac{X}{p} - \frac{1-X}{1-p}\right)^2\right] = E\left[\left(\frac{X-p}{p(1-p)}\right)^2\right] = \frac{Var(X)}{p^2(1-p)^2} = \frac{1}{p(1-p)}.$$

So, $\bar{X} \approx N(p, p(1-p)/n) \equiv (\bar{X}-p)/\sqrt{p(1-p)/n} \approx N(0,1)$. Then

$$P\left(z_{\alpha/2} \le \frac{\bar{X}-p}{\sqrt{p(1-p)/n}} \le z_{1-\alpha/2}\right) \approx \Phi\big(z_{1-\alpha/2}\big) - \Phi\big(z_{\alpha/2}\big) = 1-\alpha$$
$$\Rightarrow P\left(-z_{1-\alpha/2} \le \frac{\bar{X}-p}{\sqrt{p(1-p)/n}} \le z_{1-\alpha/2}\right) = P\left(\frac{(\bar{X}-p)^2}{(p-p^2)/n} \le z_{1-\alpha/2}^2\right) \approx 1-\alpha$$
$$\Rightarrow P\left(\left(1+\frac{z^2}{n}\right)p^2 - \left(2\bar{X} + \frac{z^2}{n}\right)p + \bar{X}^2 \le 0\right) \approx 1-\alpha, \text{where } z = z_{1-\alpha/2}$$

Solutions of the quadratic equation $(1+z^2/n)p^2 - (2\bar{X} + z^2/n)p + \bar{X}^2 = 0$ are:

$$p = \frac{\left(2\bar{X} + \frac{z^2}{n}\right) \pm \sqrt{\left(2\bar{X} + \frac{z^2}{n}\right)^2 - 4\left(1+\frac{z^2}{n}\right)\bar{X}^2}}{2(1+z^2/n)} = \frac{\left(2\bar{X} + \frac{z^2}{n}\right) \pm \sqrt{\frac{4z^2\bar{X}}{n} + \frac{z^4}{n^2} - \frac{4z^2\bar{X}^2}{n}}}{2(1+z^2/n)}$$
$$= \frac{(\bar{X} + z^2/2n) \pm \big(z/\sqrt{n}\big)\sqrt{\bar{X} - \bar{X}^2 + z^2/4n}}{1+z^2/n} \approx \bar{X} \pm \frac{z}{\sqrt{n}}\sqrt{\bar{X}(1-\bar{X})},$$

as z^2/n, $z^2/2n$, $z^2/4n$ are insignificant compared to 1, $\bar{X}$, $\bar{X}(1-\bar{X})$ respectively for large n. Note that both the solutions are real. From the theory of quadratic equations, if p is within these two solutions, then $(1+z^2/n)p^2 - (2\bar{X} + z^2/n)p + \bar{X}^2 \le 0$ and vice-versa. Then

$$P\left(\bar{X} - z_{1-\alpha/2}\sqrt{\bar{X}(1-\bar{X})/n} \le p \le \bar{X} + z_{1-\alpha/2}\sqrt{\bar{X}(1-\bar{X})/n}\right) \approx 1-\alpha$$

So, for large n, $\left[\bar{X} - z_{1-\alpha/2}\sqrt{\bar{X}(1-\bar{X})/n}\,, \bar{X} + z_{1-\alpha/2}\sqrt{\bar{X}(1-\bar{X})/n}\right]$ is a $(1-\alpha) \times 100\%$ 'approximate' confidence interval for p of $Ber(p)$. For any other quantity, we can obtain an exact/approximate confidence interval following the above approaches.

Introduction to hypothesis testing

Hypothesis testing is about accepting/rejecting an assertion based on data. Let us consider the claim that a coin is not fair. To verify the claim, we toss the coin some number of times and note the outcomes. If the numbers of heads and tails differ sufficiently, then we accept the claim, else we reject it. However, we can never be sure whether the claim is correct or not, even if we get all heads or all tails or something else. Another concern is that how much of difference in the numbers of heads and tails can be regarded as sufficient. Our decision in this context, whatever it may be, is prone to errors. We try to control the errors in a systematic manner. This requires some formalism, as described below.

Since we cannot prove or disprove the claim like we do in Mathematics, we shall not make any direct comment in favor or against the claim. Instead, we consider opposite of the claim, referred to as the null hypothesis, and see if we can reject the null hypothesis based on data. Rejection of the null hypothesis is indirect acceptance of the original claim, referred to as the alternate hypothesis, though we do not say so formally. On the other hand, if we can't reject the null hypothesis, then we cannot accept the alternate hypothesis. In the above example, the null and alternate hypothesis are $H_0: p = 0.5$ and $H_1: p \neq 0.5$, where the underlying random variable is $X \sim Ber(p)$. With this formalism, we can make two types of error in hypothesis testing, as described in the table below.

Decision	H_0 is true	H_0 is false
Reject H_0	Type-I error	No error
Do not reject H_0	No error	Type-II error

Type-I error is about rejecting H_0 (i.e., accepting H_1) incorrectly and Type-II error is about not rejecting H_0 (i.e., not accepting H_1) incorrectly. Unfortunately, we cannot reduce both the errors simultaneously. Let S denote the set of all realizations of the random sample. Then our rule for rejecting H_0, whatever it may be, partitions S into the set of realizations for which we reject H_0, denoted by R, and the set of realizations for which we do not reject H_0, denoted by $R^c = S \setminus R$. With this, we can calculate error probabilities as:

$$P(Type\ I\ error) = P(Reject\ H_0|H_0\ is\ true) = P(R|H_0\ is\ true)$$
$$P(Type\ II\ error) = P(Do\ not\ reject\ H_0|H_0\ is\ false) = P(R^c|H_0\ is\ false)$$
$$= 1 - P(R|H_0\ is\ false)$$

Observe that the only way of reducing Type-I error is to shrink R. This process also reduces $P(R|H_0\ is\ false)$, and thus increases Type-II error. Since we cannot reduce both the errors simultaneously, we balance them by fixing Type-I error at some pre-decided level and then minimizing Type-II error. This approach controls Type-I more strongly than Type-II error. Type-I error is about incorrectly rejecting H_0, i.e., accepting the original claim when it is false, whereas Type-II error is about not rejecting H_0 incorrectly, i.e., not accepting the claim when it is true. We are conservative about accepting the claim because we do not want to see the claim getting refuted in future, even if that means increased Type-II error. However, we do not want to unnecessarily increase Type-II error. Therefore, we minimize it while keeping Type-I error within a pre-decided level.

Based on the above principle, hypothesis testing is essentially about partitioning the set of realizations S into R and R^c such that $P(Type\ II\ error) = P(R^c|H_0\ is\ false)$ is minimized while $P(Type\ I\ error) = P(R|H_0\ is\ true) \leq \alpha$, the pre-decided threshold. Once we decide R, we reject H_0 if the realization is in R, otherwise we do not reject H_0. Let us understand this process through the claim that a coin is not fair, i.e., $H_0: p = 0.5$ vs. $H_1: p \neq 0.5$. Assume that the coin was tossed thrice. Then $S = \{111,110,101,011,100,010,001,000\}$, where 1 denotes head and 0 denotes tail. Let us consider the Type-I error threshold $\alpha = 0.25$. We have $2^{|S|} = 2^8$ subsets of S, which is same as the number of choices for R. We have to identify those choices that meet Type-I error bound, and then from among them, pick the one with the least Type-II error. Let us consider the choice $R = \{111,000\}$. Then

$$P(Type\ I\ error) = P(R|H_0\ is\ true) = P(\{111,000\}|p = 0.5) = 0.5^3 + 0.5^3 = 0.25 \leq \alpha$$
$$\begin{aligned} P(Type\ II\ error) &= P(R^c|H_0\ is\ false) = 1 - P(R|p \neq 0.5) \\ &= 1 - \int_0^1 P(\{111,000\}|p = \theta, p \neq 0.5) P(p = \theta|p \neq 0.5) \\ &= 1 - \int_0^1 \{\theta^3 + (1-\theta)^3\} f_{p|p\neq 0.5}(\theta) d\theta \end{aligned}$$

We do not know $f_{p|p\neq 0.5}(\theta)$, which captures the likelihood of $p = \theta$ given that H_0 is false. Thus, we cannot obtain the numerical value of Type-II error. Since $f_{p|p\neq 0.5}(\theta)$ does not change with R, focusing on the other part of the integrand, which is $\theta^3 + (1-\theta)^3$ for $R = \{111, 000\}$, helps us minimize Type-II error. This is approach is known as robust optimization. There are other approaches, and one such is to assume a form for $f_{p|p\neq 0.5}(\theta)$. Let us assume that all values of p are equally likely under false H_0, i.e., $f_{p|p\neq 0.5}(\theta) = 1$. Then

$$P(Type\ II\ error) = 1 - \int_0^1 \{\theta^3 + (1-\theta)^3\} d\theta = 1 - 2\left[\frac{\theta^4}{4}\right]_0^1 = 0.5$$

In the table below, we note down those possibilities for R that meets the Type-I error bound and mention the corresponding error probabilities.

R	$P(Type\ I)$	$P(Type\ II)$
$\emptyset$	0	1
$\{111\}$ or $\{000\}$	0.125	0.75
$\{110\}$ or $\{101\}$ or $\{011\}$ or $\{100\}$ or $\{010\}$ or $\{001\}$	0.125	11/12

{111,000}	0.25	0.5
{111,110} or {111,101} or {111,011} or {111,100} or {111,010} or {111,001} or {110,000} or {101,000} or {011,000} or {100,000} or {010,000} or {001,000}	0.25	$2/3$
{110,101} or {110,011} or {101,011} or {110,100} or {110,010} or {110,001} or {101,100} or {101,010} or {101,001} or {011,100} or {011,010} or {011,001} or {100,001} or {100,010} or {010,001}	0.25	$5/6$

$R = \{111,000\}$ minimizes Type-II error while ensuring that Type-I error is bounded by $\alpha = 0.25$. Therefore, the best rejection set is $\{111,000\}$, i.e., we shall reject $H_0: p = 0.5$ if we get three heads or three tails in the coin tosses. This kind of optimization, however, is impossible in most situations, including this one if the coin is tossed large number of times. So, we settle with a reasonably good solution (i.e., a solution that meets the Type-I error bound and offers a reasonably small value of Type-II error), if it is easy to obtain. There are several approaches for obtaining a good solution, and one such approach relies on confidence interval.

Consider the previous hypothesis, i.e., $H_0: p = 0.5$ and $H_1: p \neq 0.5$, where the underlying random variable is $X \sim Ber(p)$. Let $X_1, X_2, \dots, X_n$ be a simple random sample of X and n is large. Then $\left[\bar{X} - z_{1-\alpha/2}\sqrt{\bar{X}(1-\bar{X})/n}, \bar{X} + z_{1-\alpha/2}\sqrt{\bar{X}(1-\bar{X})/n}\right]$, to be denoted by $[L, U]$, is a $(1-\alpha) \times 100\%$ confidence interval for p, i.e., $P(p \in [L, U]) = 1 - \alpha \equiv P(p \notin [L, U]) = \alpha$ for the unknown quantity p which can take any value in $(0,1)$. Let us consider the rule 'Reject H_0 if $0.5 \notin [L, U]$'. Note that the rule implicitly partitions S into $R = \{s \in S: 0.5 \notin [L(s), U(s)]\}$ and $R^c = S \setminus R$. For this rule,

$$P(Type\ I\ error) = P(Reject\ H_0 | H_0\ is\ true) = P(0.5 \notin [L, U] | p = 0.5)$$
$$= P(p \notin [L, U] | p = 0.5) = \alpha, \text{as } P(p \notin [L, U]) = \alpha\ \forall p \in (0,1).$$
$$P(Type\ II\ error) = P(Do\ not\ reject\ H_0 | H_0\ is\ false) = P(0.5 \in [L, U] | p \neq 0.5)$$

The rule meets Type-I error bound, but we cannot calculate Type-II error. We can hope it to be reasonably small. In general, for $X \sim Ber(p)$, if we want to test $H_0: p = p_0$ vs. $H_1: p \neq p_0$, a valid rule would be to reject H_0 if $p_0 \notin [L, U]$, where $[L, U]$ denotes a $(1-\alpha) \times 100\%$ confidence interval for p. Following this principle, we can obtain valid rules for hypothesis about other quantities of interest.

Consider the hypothesis $H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$, where the underlying random variable $X \sim N(\mu, \sigma^2)$. Let $X_1, X_2, \dots, X_n$ be a simple random sample of X. Then $\left[\bar{X} - t_{1-\alpha/2}(S/\sqrt{n}), \bar{X} + t_{1-\alpha/2}(S/\sqrt{n})\right]$, to be denoted by $[L, U]$, is a $(1-\alpha) \times 100\%$ confidence interval for μ. Then following the previous arguments, 'Reject H_0 if $\mu_0 \notin [L, U]$' is a valid rule. Consider a different hypothesis $H_0: \mu \leq \mu_0$ vs. $H_1: \mu > \mu_0$. Here, we consider a $(1-\alpha) \times 100\%$ left-sided confidence interval $\left[\bar{X} - t_{1-\alpha}(S/\sqrt{n}), \infty\right)$, to be denoted by $[LB, \infty)$. Then $P(\mu \in [LB,$

$\infty)\big) = 1-\alpha \equiv P(\mu < LB) = \alpha$ for unknown μ which can take any value in $\mathbb{R}$. Consider the rule 'Reject H_0 if $\mu_0 < LB$'. For this rule,

$$P(Type\ I\ error) = P(Reject\ H_0|H_0\ is\ true) = P(\mu_0 < LB|\mu \le \mu_0)$$

If $\mu = \mu_0$, then $P(Type\ I\ error) = P(\mu_0 < LB|\mu = \mu_0) = P(\mu < LB|\mu = \mu_0) = \alpha$

If $\mu = \mu' < \mu_0$, $P(Type\ I\ error) = P(\mu_0 < LB|\mu = \mu') = P(\mu' < LB - (\mu_0 - \mu')|\mu = \mu')$

$$\le P(\mu' < LB|\mu = \mu') = P(\mu < LB|\mu = \mu') = \alpha$$

The rule meets Type-I error bound under all possible values of μ under H_0. Therefore, it is a valid rule. Now, consider the hypothesis $H_0: \mu \ge \mu_0$ vs. $H_1: \mu < \mu_0$. Here, we consider right-sided confidence interval $\big(-\infty, \bar{X} + t_{1-\alpha}(S/\sqrt{n})\big]$, to be denoted by $(-\infty, UB]$. Following the above arguments, 'Reject H_0 if $\mu_0 > UB$' is a valid rule. *Show that it meets the Type-I error bound*. Following the discussion so far and the confidence intervals that we obtained earlier, we can propose the following valid rules for certain hypothesis.

Randomness	Null hypothesis	Rule: Reject H_0 if
$N(\mu, \sigma^2)$	$H_0: \mu = \mu_0$	$\mu_0 \notin \big[\bar{X} - t_{1-\alpha/2}(S/\sqrt{n}), \bar{X} + t_{1-\alpha/2}(S/\sqrt{n})\big]$ $\equiv (\bar{X} - \mu_0)/(S/\sqrt{n}) \notin \big[-t_{1-\alpha/2}, t_{1-\alpha/2}\big]$
$N(\mu, \sigma^2)$	$H_0: \mu \le \mu_0$	$\mu_0 < \bar{X} - t_{1-\alpha}(S/\sqrt{n}) \equiv (\bar{X} - \mu_0)/(S/\sqrt{n}) > t_{1-\alpha}$
$N(\mu, \sigma^2)$	$H_0: \mu \ge \mu_0$	$\mu_0 > \bar{X} + t_{1-\alpha}(S/\sqrt{n}) \equiv (\bar{X} - \mu_0)/(S/\sqrt{n}) < -t_{1-\alpha}$
$N(\mu, \sigma^2)$	$H_0: \sigma^2 = \sigma_0^2$	$\sigma_0^2 \notin \big[(n-1)S^2/\chi^2_{1-\alpha/2}, (n-1)S^2/\chi^2_{\alpha/2}\big]$ $\equiv (n-1)S^2/\sigma_0^2 \notin \big[\chi^2_{\alpha/2}, \chi^2_{1-\alpha/2}\big]$
$N(\mu, \sigma^2)$	$H_0: \sigma^2 \le \sigma_0^2$	$\sigma_0^2 < (n-1)S^2/\chi^2_{1-\alpha} \equiv (n-1)S^2/\sigma_0^2 > \chi^2_{1-\alpha}$
$N(\mu, \sigma^2)$	$H_0: \sigma^2 \ge \sigma_0^2$	$\sigma_0^2 > (n-1)S^2/\chi^2_{\alpha} \equiv (n-1)S^2/\sigma_0^2 < \chi^2_{\alpha}$
$Exp(\lambda)$	$H_0: \lambda = \lambda_0$	$\lambda_0 \notin \big[\chi^2_{\alpha/2}/2n\bar{X}, \chi^2_{1-\alpha/2}/2n\bar{X}\big]$ $\equiv 2n\bar{X}\lambda_0 \notin \big[\chi^2_{\alpha/2}, \chi^2_{1-\alpha/2}\big]$
$Exp(\lambda)$	$H_0: \lambda \le \lambda_0$	$\lambda_0 < \chi^2_{\alpha}/2n\bar{X} \equiv 2n\bar{X}\lambda_0 < \chi^2_{\alpha}$
$Exp(\lambda)$	$H_0: \lambda \ge \lambda_0$	$\lambda_0 > \chi^2_{1-\alpha}/2n\bar{X} \equiv 2n\bar{X}\lambda_0 > \chi^2_{1-\alpha}$
$Pois(\mu)$, large n	$H_0: \mu = \mu_0$	$\mu_0 \notin \Big[\bar{X} - z_{1-\alpha/2}\sqrt{\bar{X}/n}, \bar{X} + z_{1-\alpha/2}\sqrt{\bar{X}/n}\Big]$ $\equiv (\bar{X} - \mu_0)/\sqrt{\bar{X}/n} \notin \big[-z_{1-\alpha/2}, z_{1-\alpha/2}\big]$
$Ber(p)$, large n	$H_0: p = p_0$	$p_0 \notin \Big[\bar{X} - z_{1-\alpha/2}\sqrt{\bar{X}(1-\bar{X})/n}, \bar{X} + z_{1-\alpha/2}\sqrt{\bar{X}(1-\bar{X})/n}\Big]$ $\equiv (\bar{X} - p_0)/\sqrt{\bar{X}(1-\bar{X})/n} \notin \big[-z_{1-\alpha/2}, z_{1-\alpha/2}\big]$

Note that all the above rules involve strict inequality. If these are replaced by $\le$ or $\ge$, nothing changes as the underlying point estimator (leading to the confidence intervals) are continuous or near-continuous due to large n. We can construct confidence intervals for more quantities of interest and obtain valid rules for hypothesis testing. There is lot more in hypothesis testing than this introduction. We have discussed only the core idea and one of the methods. Before we conclude, we will mention about a quantity called p-value.

Consider $H_0: \mu \leq 100$ vs. $H_1: \mu > 100$ for $N(\mu, \sigma^2)$. Let 116,118,101,106,108,104,96,96, 93,102 be the realization of a simple random sample of $N(\mu, \sigma^2)$. Then $(\bar{X} - \mu_0)/(S/\sqrt{n}) =$ 1.52. Let the Type-I error bound be $\alpha = 0.1$. Then $t_{1-\alpha} = 1.38$. Since $(\bar{X} - \mu_0)/(S/\sqrt{n}) >$ $t_{1-\alpha}$, we reject H_0. Consider another realization of the simple random sample 106,84,117, 101,106,108,122,119,111,110. In this case, $(\bar{X} - \mu_0)/(S/\sqrt{n}) = 2.47$. Again, we reject H_0 as $(\bar{X} - \mu_0)/(S/\sqrt{n}) > t_{1-\alpha}$. If you observe closely, value of as $(\bar{X} - \mu_0)/(S/\sqrt{n})$ is farther away from the threshold $t_{1-\alpha}$ in the second case. In this sense, rejection of H_0 is stronger in the second case. We capture this aspect of hypothesis testing by p-value.

p-value is defined as the smallest value of Type-I error bound for which H_0 can be rejected. In the first of the above cases, $1 - F_{T(9)}(1.52) = 0.08$. So, H_0 could be rejected for all $\alpha \geq$ 0.08, and thus, the p-value is 0.08. In the second case, $1 - F_{T(9)}(1.52) = 0.02$. So, H_0 could be rejected for all $\alpha \geq 0.02$, and hence, the p-value is 0.02. Since p-value is smaller in the second case, rejection of H_0 is stronger. p-value must be at most the Type-I error bound α for rejection of H_0. Now, consider the hypothesis $H_0: \mu \geq 100$ vs. $H_1: \mu < 100$ for $N(\mu, \sigma^2)$ and the realization 105,104,87,98,88,97,83,106,93,93. *Verify that H_0 can be rejected at $\alpha = 0.1$ and the p-value is 0.05.* Finally, consider the hypothesis $H_0: \mu = 100$ vs. $H_1: \mu \neq 100$ and two realizations 109,68,71,106,99,86,125,72,84,94 and 116,97,110,132,113,103,76,113, 121,112. Verify that in the first case, H_0 cannot be rejected at $\alpha = 0.1$ and the p-value is 0.18, and in the second case, H_0 can be rejected and the p-value is 0.08.

Practice problems

Book-1: A Modern Introduction to Probability and Statistics by Dekking et al.

Confidence interval for μ
Book-1, Chapter-23, Exercise No. 1, 3, 6, 8, 10

Other confidence intervals
Book-1, Chapter-24, Exercise No. 1, 3, 6, 9, 10

Hypothesis testing
Book-1, Chapter-26, Exercise No. 1, 3, 4
Book-1, Chapter-27, Exercise No. 1, 3, 4