Testing Statistical Hypothesis.

Let $X$ be a r.v. with pdf $f(x\mid \theta)$. The problem is to test the null hypothesis

$H_0: \theta \in \Omega_0$

against the alternative

$H_1: \theta \in \Omega_1$

where $\Omega_0 \cap \Omega_1 = \emptyset$ and $\Omega_0 \cup \Omega_1 = \mathbb{R}$ be the parameter space as specified by the decision maker.

The claim made by the decision maker is taken to be the alternative hypothesis.

If $\Omega_0$ completely specifies the distribution we call $H_0$ simple if it does not, we call $H_0$ composite.

Example: $X \sim N(\mu, \sigma^2)$ $H_0: \mu = 3$, $H_1: \mu > 3$

$H_0$ is simple if $\sigma^2$ is known, $H_0$ is composite if $\sigma^2$ is unknown. $H_1$ is composite.

True but unknown state of nature

<table>
<thead>
<tr>
<th>Decision</th>
<th>Reject $H_0$</th>
<th>$H_0$ if false</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Type I Error</td>
<td>Correct Decision</td>
</tr>
<tr>
<td></td>
<td>Correct Decision</td>
<td>Type II Error</td>
</tr>
</tbody>
</table>

Let $X = (X_1, \ldots, X_n)$ be a r.v. of size $n$. A test of $H_0$ against $H_1$ is a division of the sample space into the regions $C$ and $C^c$. If $X \in C$ is observed then we reject $H_0$. $C$ is called the critical (rejection) region.
Choice of the critical region \( C \) is to be such that

\[
\alpha(\theta) = P[ \text{Committing a Type I Error} ] = P[ \text{Rejecting } H_0 \mid H_0 \text{ is true} ]
= P[ X \in C \mid \theta \in \theta_0 ] \leq \alpha \quad \text{i.e. } \alpha = \sup_{\theta \in \theta_0} \alpha(\theta)
\]

and

\[
1 - \beta(\theta) = 1 - P[ \text{Committing a Type II Error} ]
= 1 - P[ \text{Accepting } H_0 \mid H_0 \text{ is false} ]
= P[ \text{Rejecting } H_0 \mid H_0 \text{ is false} ]
= P[ X \in C \mid \theta \in \theta_1 ] \text{ is large.}
\]

\( \alpha \) is called the significance level of the test or the size of the critical region.

\[
\gamma(\theta) = P[ \text{Rejecting } H_0 ] = P[ X \in C ] = \left\{ \begin{array}{ll} \alpha(\theta) & \theta \in \theta_0 \\ 1 - \beta(\theta) & \theta \in \theta_1 \end{array} \right.
\]

is called the power function.

**Question:** How do we find \( \alpha \)?

**Example 1:** \( N(\theta, 100) \)

\( H_0: \theta \leq 75 \)

\( H_1: \theta > 75 \)

**Test** 1: Suppose we decide to test \( H_0 \) against \( H_1 \) by taking a \( \sigma^2 \) of size \( n = 25 \).
Suppose we set
\[ G = \{ (x_1, \ldots, x_{25}) : X > 75 \} \]
Then, the power function of this test is
\[ \gamma_1(\theta) = P[X \in G] = P[X > 75] = P[Z > \frac{75 - \theta}{\sigma/\sqrt{25}}] = P[Z > \frac{75 - \theta}{\sigma/2}] \]
where \( Z \sim \mathcal{N}(0, 1) \).

\[ \alpha = \sup_{\theta \in \Theta_0} \gamma_1(\theta) = 0.5 \quad \text{which is too high.} \]

Test 1 has undesirable features. Let's try the device a test for which \( \alpha \) is small.

**Test 2.** \( n = 25 \)
\[ G = \{ (x_1, \ldots, x_{25}) : X > 78 \} \]
\[ \gamma_2(\theta) = P[X > 78] = P[Z > \frac{78 - \theta}{\sigma/2}] \]
\[ \delta_2(72) = 0.006 \]
\[ \delta_2(75) = 0.067 \]
\[ \delta_2(77) = 0.209 \]
\[ \delta_2(79) = 0.691 \]

\( \alpha = 0.067 \) is satisfactory, but
\[ \gamma_2(77) = 0.309 \quad \gamma_2(79) = 0.691 \] are not large.

Test 2 has also undesirable features.
Test 3: Let's first select a power function 
\( Y_j(\theta) \) such that 
\[ Y_j(75) = 0.159 \quad \text{(Small)} \]
\[ Y_j(77) = 0.841 \quad \text{(Large)} \]

Let \( \mathcal{G} = \{ x_1, \ldots, x_n \} : x > c \) \(\forall j \) for some \( n \) and \( c \).

\[ Y_j(\theta) = P \left( \frac{\bar{x} - c}{\sqrt{\sigma / n}} \right) \]

\[ Y_j(75) = 0.159 \quad \Rightarrow \quad \frac{c - 75}{10 / \sqrt{n}} = 1 \quad \{ \text{Solve for } n \} \]

\[ Y_j(77) = 0.841 \quad \Rightarrow \quad \frac{c - 77}{10 / \sqrt{n}} = -1 \]

We get \( n = 100 \), \( c = 76 \).

Test 3 has more desirable power function than those of Tests 1 and 2, but it requires a larger sample size (\( n = 100 \)) than the earlier tests, which approximated \( n = 25 \).

**Additional Comments:**

\( N(\theta, \sigma^2) \) \( \sigma^2 \) is known. \( \theta_0 \) is a given constant.

\( H_0: \theta = \theta_0 \)

\( H_1: \theta > \theta_0 \)

\( \omega = \{ \theta_0 \}: \theta > \theta_0 \}

\( \omega_1 = \{ \theta_0 \}: \theta < \theta_0 \}

Support \( \mathcal{G} = \{ x: x > c \} \)
if we want a test of size \( \alpha \) then

\[
\alpha = P \left( X > c \mid \theta = \theta_0 \right) = P \left( \frac{X - \theta_0}{\sigma / \sqrt{n}} > \frac{c - \theta_0}{\sigma / \sqrt{n}} \mid \theta = \theta_0 \right) = P \left( Z > \frac{c - \theta_0}{\sigma / \sqrt{n}} \right)
\]

where \( Z = \frac{X - \theta_0}{\sigma / \sqrt{n}} \sim N(0,1) \) under \( H_0 \).

\[
c = \theta_0 + \frac{z_{\alpha}}{\sqrt{n}}
\]

\[
\bar{X} \sim \frac{X - \theta_0}{\sigma / \sqrt{n}} \Rightarrow Z \sim \frac{\bar{X} - \theta_0}{\sigma / \sqrt{n}}
\]

\[
H_0: \theta = \theta_0
\]

\[
H_1: \theta \neq \theta_0
\]

Suppose \( \bar{X} \sim X \): \( \bar{X} < c_1 \), or \( \bar{X} \geq c_2 \)

if we want a test of size \( \alpha \) then

\[
\alpha = P \left( X < c_1 \mid \theta = \theta_0 \right) + P \left( X \geq c_2 \mid \theta = \theta_0 \right) = P \left( Z < \frac{c_1 - \theta_0}{\sigma / \sqrt{n}} \right) + P \left( Z \geq \frac{c_2 - \theta_0}{\sigma / \sqrt{n}} \right)
\]

\[
\alpha = \frac{z_{\alpha}}{\sqrt{n}} + \frac{z_{1 - \alpha}}{\sqrt{n}}
\]

\[
c_1 = \theta_0 + \frac{z_{\alpha}}{\sqrt{n}}
\]

\[
c_2 = \theta_0 + \frac{z_{1 - \alpha}}{\sqrt{n}}
\]

\[
\bar{X} \leq \frac{c_2 - \theta_0}{\sigma / \sqrt{n}} \quad \text{or} \quad \bar{X} \geq \frac{c_1 - \theta_0}{\sigma / \sqrt{n}}
\]
Acceptance region: 

\[ AR \supset C \]

Confidence Interval: 

\[ CI \supset \{ \theta \in \Theta : x - \frac{z}{2} \frac{s}{\sqrt{n}} < \theta < x + \frac{z}{2} \frac{s}{\sqrt{n}} \} \]

If \( \theta_0 \in CI \), then \( H_0 \) will be accepted.

**Theorem:** For each \( \theta_0 \in \Theta \), let \( AR(\theta_0) \) be the acceptance region at a level of test \( \alpha \); \( \theta_0 = \theta_0 \). For each \( x \in \mathbb{R} \), define the set \( CI(x) \) in the parameter space by

\[ CI(x) = \{ \theta \in \Theta : x \in AR(\theta) \} \]

Then the random set \( CI(x) \) is a \( 1 - \alpha \) confidence set.

Conversely, let \( CI(x) \) be a \( 1 - \alpha \) confidence set. For any \( \theta_0 \in \Theta \), define

\[ AR(\theta_0) = \{ x : \theta_0 \in CI(x) \} \]

Then \( AR(\theta_0) \) is the acceptance region of a level \( \alpha \) test of \( H_0 : \theta = \theta_0 \).

**P-Value:** If \( T(x) \) is the statistic which is used to test \( H_0 \) and if the critical region is of the form \( T(x) \geq c \), then

\[ P \text{-value} = P \left[ T(x) \geq T_{obs} \mid H_0 \right] \]

If the critical region is of the form \( T(x) \leq c \), then

\[ P \text{-value} = P \left[ T(x) \leq T_{obs} \mid H_0 \right] \]

where \( T_{obs} \) is the observed value of \( T(x) \) for the given sample.

Suppose \( \alpha = P \left[ T(x) \geq c \mid H_0 \right] \)

\[ P \text{-value} = P \left[ T(x) \geq T_{obs} \mid H_0 \right] \]

\( H_0 \) is rejected at level of significance \( \alpha \).

If we change an \( \alpha \) such that \( \alpha > P \text{-value} \).
If the critical region is of the form
given 
\( P\)-value = \( 2P\left[ T(x) \leq T_{0.025} | \text{Ho} \right] \)
\( P\)-value = \( 2P\left[ T(x) \geq T_{0.025} | \text{Ho} \right] \)

whichever is less than one.

**Example.** (Test for \( \mu \) under Normality)
\( X \sim N(\mu, \sigma^2) \)
\( \text{Ho: } \mu = \mu_0, \quad \mu_0 \) is a given constant.
\( \sigma^2 \) is unknown.

Test statistic \( T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t(n-1) \) under Ho.

For a given \( \alpha \), we reject Ho if \( T_{obs} \geq t_{\alpha, n-1} \).

\( P\)-value = \( P\left[ T(n-1) \text{ or } > T_{obs} \right] \).

**Example.** (Large sample test for \( \mu \)).
Let \( X_1, \ldots, X_n \) be a r.s. from a distribution with mean \( \mu \) and variance \( \sigma^2 \).
\( \text{Ho: } \mu = \mu_0, \quad \mu_0 \) is a given constant.
\( \sigma^2 \) is unknown.

Test statistic \( Z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim N(0,1) \) under Ho.

For a given \( \alpha \), reject Ho if \( Z_{obs} \geq Z_{\alpha} \).

\( P\)-value = \( P\left[ N(0,1) \text{ or } > Z_{obs} \right] \).
Example (Randomized Test).

Let $X_i \sim X_{10}$ be a r.v. from $\text{Poisson}(\theta)$. 

$H_0$: $\theta = 0.1$

$H_1$: $\theta > 0.1$

Let $Y = \frac{10}{n} \sum X_i$, $Y \sim \text{Poisson}(10\theta)$. 

Suppose we reject $H_0$ when $Y > 3$. Then

$\alpha = P(Y > 3 | \theta = 0.1) = 1 - \sum_{y=0}^{2} \frac{e^{-1} 1^y}{y!} = 0.08$

If the critical region defined by $Y > 4$ is used, then

$\alpha = P(Y > 4 | \theta = 0.1) = 1 - \sum_{y=0}^{3} \frac{e^{-1} 1^y}{y!} = 0.019$

If a level of significance of $\alpha = 0.05$ is desired, then it can be achieved as follows:

Let $W \sim \text{Bernoulli}(p)$. $W$ is selected independently of the sample.

Reject $H_0$ if $Y > 4$ or if $Y = 3$ and $W = 1$.

$\alpha = 0.05 = P(Y > 4 | \theta = 0.1) + P(Y = 3, W = 1 | \theta = 0.01)$

$= 0.019 + P(W = 1) P(Y = 3 | \theta = 0.01)$

$= 0.019 + P\left(\frac{e^{-1}}{3^1}\right) = 0.019 + P(0.061) \rightarrow p = 0.508$

Thus we let $W \sim \text{Bernoulli}(p = 0.508)$.

Reject $H_0$ if $Y > 4$ or if $Y = 3$ and $W = 1$. Then $\alpha = 0.05.$
Chi-Square Tests

Let \( X_i \sim \text{Binomial}(n, p_i) \)

\[
Y = \frac{X_i - np_i}{\sqrt{np_i(1-p_i)}} \xrightarrow{D} N(0, 1) \text{ by CLT.}
\]

Question: What is the limiting distribution of \( Q = Y^2 \) as \( n \to \infty \)?

Let \( G_n(\cdot) \) denote the cdf of \( Y \)

\[
H_n(\cdot) = \Phi(\cdot) \quad \text{and} \quad \Phi(\cdot) = N(0, 1).
\]

\[
H_n(q) = P(Q \leq q) = P(Y^2 \leq q) = P(-\sqrt{q} \leq Y \leq \sqrt{q})
\]

\[
= G_n(\sqrt{q}) - G_n(-\sqrt{q})
\]

\[
\lim_{n \to \infty} H_n(q) = \lim_{n \to \infty} G_n(\sqrt{q}) - \lim_{n \to \infty} G_n(-\sqrt{q})
\]

\[
= \Phi(\sqrt{q}) - \Phi(-\sqrt{q})
\]

\[
= \int_{-\sqrt{q}}^{\sqrt{q}} \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} dw
\]

\[
= 2 \int_{0}^{\sqrt{q}} \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} dw
\]

Let \( w = w^2 \)

\[
= \int_{0}^{\sqrt{q}} \frac{1}{\Gamma(\frac{1}{2}) 2^{1/2}} w^{\frac{1}{2}-1} e^{-w/2} dw
\]

Not: \( \Phi(\frac{1}{2}) = \frac{1}{\sqrt{\pi}} \).

Provided \( q > 0 \):

If \( q < 0 \) then \( \lim_{n \to \infty} H_n(q) = 0 \).

Thus limiting distribution of \( Q = Y^2 \) is \( \chi^2 \).

\[
\frac{(X_i - np_i)^2}{np_i(1-p_i)} \xrightarrow{D} \chi^2_{(1)}
\]
Let \( X_2 = n - X_1 \), \( P_2 = 1 - P_1 \)

\[
Q_1 = \frac{(X_1 - nP_1)^2}{nP_1(1-P_1)} = \frac{(X_1 - nP_1)^2}{nP_1} + \frac{(X_1 - nP_1)^2}{n(1-P_1)}
\]

\[
= \frac{(X_1 - nP_1)^2}{nP_1} + \frac{(X_2 - nP_2)^2}{nP_2}
\]

\[
\Rightarrow Q_1 = \sum_{i=1}^{2} \frac{(X_i - nP_i)^2}{nP_i} \xrightarrow{L} X_{(1)}^2
\]

**Theorem:** if \((X_1, \ldots, X_k) \sim \text{Multinomial}(n; P_1, \ldots, P_k)\)

Then \(Q_{k-1} = \sum_{i=1}^{k} \frac{(X_i - nP_i)^2}{nP_i} \xrightarrow{L} X_{(k-1)}^2\)

where \(X_k = n - X_1 - \cdots - X_{k-1}\), \(P_k = 1 - P_1 - \cdots - P_{k-1}\)

Chi-Square Test: Suppose \((X_1, \ldots, X_k) \sim \text{Multinomial}(n; P_1, \ldots, P_k)\)

\[H_0: P_1 = \pi_0, \ldots, P_k = \pi_0\quad \sum_{i=1}^{k} \pi_i = 1, \pi_0 > 0\]

\[H_1: \text{At least one inequality}\]

\[Q_{k-1} = \sum_{i=1}^{k} \frac{(X_i - nP_i)^2}{nP_i} \sim X_{(k-1)}^2\] under \(H_0\) for large \(n\).

Large values of \(Q_{k-1}\) leads to rejection of \(H_0\).

\[\Rightarrow \text{Reject } H_0 \text{ if } Q_{k-1} \geq \chi^2_{\alpha, k-1}\]
Let $\gamma$ be a rv. Consider testing

$H_0$: $\gamma$ has a dist. with pd.f $g(y)$

$H_1$: $\gamma$ does not have a dist. with pd.f $g(y)$.

Partition the space of $\gamma$ into $k$ mutually disjoint sets $A_1\ldots A_k$.

Let $\pi_i = P(Y \in A_i)$, $h_i = \int g(y)dy$ or $\frac{1}{n} \sum g(y)$.

If $g(y)$ has any unknown parameters, we need to first estimate them before we can calculate $\pi_i$'s.

Let $X_i$ be the number of $\gamma$'s in $A_i$.

Under $H_0$, $(X_1, \ldots, X_k) \sim \text{multinom}(n, \pi_1, \ldots, \pi_k)$.

Then for large $n$,

$Q = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i} \sim \chi^2_{(k-1-m)}$

where $m$: Number of parameters estimated.

---

Suppose $(X_{i1}, \ldots, X_{i k_i}) \sim \text{multinom}(n_1, p_{i1}, p_{i2}, \ldots, p_{ik_i})$

independently

$(X_{k1}, \ldots, X_{k k_k}) \sim \text{multinom}(n_2, p_{k1}, p_{k2}, \ldots, p_{kk})$

For large $n_1$ and $n_2$

$$\sum_{i=1}^{k} \sum_{j=1}^{k_i} \frac{(X_{ij} - np_{ij})^2}{np_{ij}} \sim \chi^2_{(k_1 + k_2 - m)} = \chi^2_{(2k-2)}$$

$H_0$: \ $p_{11} = p_{k1}, \ldots, p_{k1} = p_{k2}$

$H_1$: At least one equality above.
Let $p'_i = p_{i1} = p_{i1}$ denote the common value of $p_{i1}$ and $p_{i2}$ under $H_0$.

Under $H_0$,
$$\sum_{j=1}^{k} \sum_{i=1}^{k} \frac{(x_{ij} - n_j p'_i)^2}{n_j p'_i} \sim \chi^2_{(2k-2)}$$

We need to estimate $p'_i$'s.

$$\prod_{i=1}^{k} \frac{x_{i1} + x_{i2}}{n_{i1} + n_{i2}} = \prod_{i=1}^{k} \frac{n_i!}{n_{i1}! n_{i2}!}$$

The maximum likelihood estimator of $p'_i$ is $\hat{p}'_i = \frac{x_{i1} + x_{i2}}{n_{i1} + n_{i2}}$.

Not that number of parameters estimated is $k-1$ since $\sum_{i=1}^{k} p'_i = 1$.

Then under $H_0$,
$$Q = \sum_{j=1}^{k} \sum_{i=1}^{k} \frac{(x_{ij} - n_j \hat{p}'_i)^2}{n_j \hat{p}'_i} \sim \chi^2_{(2k-2-(k-1))} \equiv \chi^2_{(k-1)}$$

Reject $H_0$ if $Q \geq X^2_{\alpha, k-1}$

Suppose $x_{ij}$'s are Multinomial $(n, \pi_{ij})$.

$$h_{ij} = x_{ij}$$

$$H_0: P_{ij} = p_i \cdot p_j \quad i = 1, \ldots, b, \quad j = 1, \ldots, b$$

$$H_1: \neq$$

$H_0$: Row and Column classifications are independent.

$H_1$: There is no association between $p_i$ row and $p_j$ column.
\[
\sum_{j=1}^{b} \sum_{i=1}^{a} \frac{(X_{ij} - n \hat{p}_{i})^2}{n \hat{p}_{i}} \sim \chi^2_{(ab-1)}
\]

Under Ho:
\[
\sum_{j=1}^{b} \sum_{i=1}^{a} \frac{(X_{ij} - n \hat{p}_{i} \hat{p}_{j})^2}{n \hat{p}_{i} \hat{p}_{j}} \sim \chi^2_{(ab-1)}
\]

M.L.E. \( \hat{p}_{i} \) is \( \frac{X_{i} \cdot}{n} \) \( i = 1 \rightarrow a \)
M.L.E. \( \hat{p}_{j} \) is \( \frac{X_{j} \cdot}{n} \) \( j = 1 \rightarrow b \)
\( \sum_{i=1}^{a} p_{i} = 1 \), \( \sum_{j=1}^{b} p_{j} = 1 \), we have estimated \((a-1) + (b-1) = a+b-2\) parameters.

Thus under Ho:
\[
Q = \sum_{j=1}^{b} \sum_{i=1}^{a} \frac{(X_{ij} - n \hat{p}_{i} \hat{p}_{j})^2}{n \hat{p}_{i} \hat{p}_{j}} \sim \chi^2_{(ab-1-(a+b-2))}
\]
\[
= \chi^2_{(a-1)(b-1)}
\]

Reject Ho if \( Q \geq \chi^2_{(a-1)(b-1)} \)

\( p \)-value = \( P[\chi^2_{(a-1)(b-1)} \geq Q_{\text{observed}}] \)

\[
Q = \sum_{j=1}^{b} \sum_{i=1}^{a} \frac{(X_{ij} - \frac{X_{ij} \cdot}{n})^2}{\frac{X_{ij} \cdot}{n}} \sim \chi^2_{(a-1)(b-1)}
\]
Generating a Random Sample.

We are concerned with generating a random sample \( X_1, \ldots, X_n \) from a given distribution.

Direct Methods: A direct method of generating a random variable is one for which there exists a closed form function \( g(u) \) such that the transformed variable \( X = g(U) \) has the desired distribution when \( U \sim \text{Uniform}(0,1) \).

Theorem: Let \( U \sim \text{Uniform}(0,1) \). Let \( F \) be a continuous distribution function. Then \( X = F^{-1}(U) \) has distribution function \( F \).

Proof: \[ P[X \leq x] = P[F^{-1}(U) \leq x] = P[U \leq F(x)] = F(x). \]

Example: Suppose \( X \sim \text{Exponential}(\beta) \)

\[ F(x) = \int_0^x \frac{1}{\beta} e^{-\frac{t}{\beta}} \, dt = 1 - e^{-\frac{x}{\beta}}, \quad x > 0 \]

then \( X = F^{-1}(U) = -\beta \ln(1-U), \quad 0 < U < 1 \).

Thus if we generate \( U_1, \ldots, U_n \) as iid uniform \((0,1)\) random variables, then \( X_i = -\beta \ln(1-U_i) \) for \( i = 1, \ldots, n \) are iid \( \text{Exponential}(\beta) \) random variables.

The relationship between and other distributions allows the generation of many random variables. If \( U_1, \ldots, U_n \) are iid \( \text{Uniform}(0,1) \) random variables, then
\[ X = -2 \sum_{i=1}^{k} \ln U_i \sim X_{2k}^2 \]
\[ X = -\beta \sum_{i=1}^{\alpha} \ln U_i \sim \Gamma(\alpha, \beta) \]
\[ X = \frac{\sum_{i=1}^{\alpha} \ln U_i}{\sqrt{\sum_{i=1}^{\alpha} \ln U_i}} \sim \text{Beta}(\alpha, \beta). \]

**Box-Muller Algorithm:**

Generate \( U_1, U_2 \) two independent \( \text{Uniform}(0,1) \) random variables. Then

\[ X_1 = \sqrt{-2 \ln U_1} \cos(2\pi U_2) \quad \text{and} \]
\[ X_2 = \sqrt{-2 \ln U_1} \sin(2\pi U_2) \]

are independent \( \mathcal{N}(0,1) \) random variables.

**The case of discrete random variables:**

Let \( X \) be a discrete r.v. taking on values
\[ x_1 < x_2 < \ldots < x_k \]
then
\[ P\left[ F(x_i) < U \leq F(x_{i+1}) \right] = F(x_{i+1}) - F(x_i) = P[X = x_{i+1}] \]

Thus to generate \( X \),

1. Generate \( U \sim \text{Uniform}(0,1) \)
2. If \( F(x_i) < U \leq F(x_{i+1}) \), set \( X = x_{i+1} \)

Define \( x_0 = -\infty \) and \( F(x_0) = 0 \).
Example. To generate a $X \sim \text{Bernoulli}(\frac{1}{3})$, generate $U \sim \text{Uniform}(0,1)$ and set

$$X = \begin{cases} 
0 & \text{if } 0 < U \leq \frac{2}{3} \\
1 & \text{if } \frac{2}{3} < U \leq 1
\end{cases}.$$ 

Example. To generate a $X \sim \text{Binomial}(n=4, \ p=\frac{5}{6})$, generate $U \sim \text{Uniform}(0,1)$ and set

$$X = \begin{cases} 
0 & \text{if } 0 < U \leq 0.20 \\
1 & \text{if } 0.20 < U \leq 0.52 \\
2 & \text{if } 0.52 < U \leq 0.81 \\
3 & \text{if } 0.81 < U \leq 0.97 \\
4 & \text{if } 0.97 < U \leq 1
\end{cases}.$$ 

Monte Carlo Integration:

We want to evaluate $\int_a^b g(x) \, dx$ where $g(x)$ is a continuous function.

$$\int_a^b g(x) \, dx = (b-a) \int_a^b g(u) \frac{1}{b-a} \, du = (b-a) \mathbb{E}[g(U)]$$

where $U \sim \text{Uniform}(a,b)$. Generate a random sample $U_1, \ldots, U_n$ from Uniform $(a,b)$, compute $Y_i = (b-a) g(U_i)$, then $\bar{Y} = (b-a) \frac{\sum_i g(U_i)}{n}$ is a consistent estimate of $\int_a^b g(x) \, dx$. 
Indirect Methods.

When no easily found direct transformation is available to generate the desired random variable, Accept-Reject Algorithm can often provide a solution.

The Accept-Reject Algorithm:

Let \( f(x) \) be a pdf. Suppose that \( Y \) is a r.v with pdf \( g(y) \) and \( U \sim \text{Uniform}(0,1) \), \( Y \) and \( U \) are independent and \( \frac{f(x)}{g(x)} \leq M, -\infty < x < \infty \) for some constant \( M < \infty \).

To generate a r.v \( X \) with pdf \( f(x) \):

a) Generate \( Y \) and \( U \)

b) If \( U \leq \frac{f(y)}{M g(y)} \), set \( X = Y \), otherwise return to step (a).

\( f(x) \) is called the target pdf, \( g(y) \) is called the instrumental or candidate density. The requirement that \( M < \infty \) can be interpreted as requiring the density of \( Y \) to have heavier tails than the density of \( X \).

Example: To generate \( X \sim \text{Beta}(2.7, 6.3) \)

a) Generate \( Y \sim \text{Beta}(2, 6) \), \( U \sim \text{Uniform}(0,1) \)

b) If \( U \leq \frac{1}{M} \frac{f(y)}{g(y)} \), set \( X = Y \), otherwise return to step (a).
For the given densities we have $M = 1.67$

**Metropolis Algorithm:**

Let $X$ have the pdf $f(x)$, $Y$ have the pdf $g(y)$ where $f$ and $g$ have common support. To generate $X$:

a) Generate $Y$, set $Z_0 = Y$.

For $i = 1, 2, \ldots$,

b) Generate $U_i \sim \text{Uniform}(0,1)$, generate $Y_i$ and calculate $P_i = \min \left\{ \frac{f(y_i)}{g(y_i)}, \frac{g(z_{i-1})}{f(z_{i-1})}, 1 \right\}$

c) Set $Z_i = \begin{cases} Y_i & \text{if } U_i \leq P_i \\ Z_{i-1} & \text{if } U_i \geq P_i \end{cases}$

Then as $i \to \infty$, $Z_i$ converges to $X$ in distribution.

In practice, after the algorithm is run for a while ($i$ gets big) then $Z_i$'s that are produced behave like variables from $f(x)$.

Metropolis algorithm is especially useful when the target density has heavy tails and it is difficult to set instrumental densities that will result in finite values of $M$ in the accept-reject algorithm.
Bootstrap Procedures:

We learn about the characteristics of a population by taking samples. As the sample represents the population, analogous characteristics of the sample should give us information about the population characteristics. The bootstrap helps us learn about the sample characteristics by taking resamples (that is, we take samples from the original sample) and use this information to infer to the population. If \( X_1, \ldots, X_n \) is the original sample, we can consider the bootstrap random variable \( X^* \)
defined as \[ P \left( X^* = X_i \right) = \frac{1}{n} \quad i=1, 2, \ldots, n \]

The bootstrap sample is simply a random sample from the distribution above, say \( X_1^*, X_2^*, \ldots, X_n^* \). This means that \( X_i^* \)'s are independent with the same distribution above.

In practice, this means that we are sampling the elements of the original sample with replacement.

\[ E(X_i^*) = \bar{x}, \quad Var(X_i^*) = \frac{1}{n} \sum (x_i - \bar{x})^2 \]

Bootstrap Confidence Intervals.

Let \( X = (X_1, \ldots, X_n) \) be the realization of a random sample drawn from a distribution with parameter \( \theta \).

Let \( \hat{\theta}(X) \) be a point estimator of \( \theta \).
1. Set $j = 1$

2. While $j \leq B$, do steps (2)-(5).

   [B is the number of bootstrap replications. In practice is often 3000 or more]

3. Let $X_j^* = (X_{j1}^*, \ldots, X_{jn}^*)$ be a random sample of size $n$ from the sample $X = (X_1, \ldots, X_n)$. That is, the observations $X_{j1}^*, \ldots, X_{jn}^*$ are drawn from $X_1, \ldots, X_n$ with replacement.

4. Let $\hat{\theta}_j^* = \hat{\theta}(X_j^*)$

5. Replace $j$ by $j+1$

6. Let $\hat{\theta}_1^* \leq \hat{\theta}_2^* \leq \ldots \leq \hat{\theta}_B^*$ denote the ordered values of $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$. Let

   $m = \lceil \frac{B}{2} \rceil$ where $\lceil \cdot \rceil$ denotes the greatest integer function. Form the interval

   $\left( \hat{\theta}_{(m)}^*, \hat{\theta}_{(B+1-m)}^* \right)$

   that is obtained 100% and $(1-\frac{1}{2})100\%$ percentiles of the sampling distribution of $\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$.

   This interval is called percentile bootstrap confidence interval for $\theta$. 
