STA 453/553 Class 1

- Introduce class and syllabus discussion

Web site: http://www.stat.unh.edu/~ghikerta/inference

Course html.

Begin with Chap. 5: Elements of Stat. Inference.

X a r.v. that follows some dist. F - not necessarily known
Sample: \( x_1, x_2 \), \( X \sim F \) \( \theta \) some family or
distributions.

Question about \( F \): mean, variance, prob?

For this course, \( \Theta \) is a parametric family, i.e.
\( \{ f(x|\theta) \colon \theta \in \Theta \} \) \( \Theta \) parameter, \( \Theta \) parameter
space \( \Theta \) an unknown quantity.

Question about \( F \) translates to a question on \( g(\theta) \),
a function of \( \theta \).

Ex: \( \{ f(x|\mu) = N(\mu, \sigma^2) \} \) \( \sigma^2 \) known - \( \mu \)?

Non-parametric situation: family \( \Theta \) may not be indexed
by parameter. Ex: \( \{ \) all cont. densities \( \}\),
\( \Theta \)=all symmetric dist. \( \).

A very important element of inference.

Random samples: see 5.1 & 5.2

\( x_1, x_2 \), \( X \) of r.v.s independent and identically
distributed (iid) each following a density \( f(x) \).

Key issue of random sampling:

\[ f(x_1, x_2, \ldots, x_n | \theta) = \prod_{i=1}^n f(x_i | \theta) \]

\[ f(x_1, x_2, \ldots, x_n | \theta) = \prod_{i=1}^n f(x_i | \theta) \]
For $k$,

$$ f(x_1 \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{1}{2\sigma^2} (x_i - \mu)^2 \right) $$

and $x_1, x_2, \ldots, x_n$ are a random sample $x_i \sim N(\mu, \sigma^2)$ \Rightarrow

$$ f(x_1, x_2, \ldots, x_n \mid \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp \left( -\frac{1}{2\sigma^2} (x_i - \mu)^2 \right) = \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right). \quad \forall$$

Obs. When the $x_i s$ are observed, $f(x_1, x_2, \ldots, x_n \mid \mu, \sigma^2)$ determines the likelihood function on $(\mu, \sigma^2)$ Chap. 6.

Random sampling refers to sampling from infinite populations. Another issue:

Samples of random samples are expressed by statistics. Def: A statistic is a function $T(x_1, x_2, \ldots, x_n)$ real- or vector-valued whose domain is the sample space of $(x_1, x_2, \ldots, x_n)$.

Sample space = set of all possible values that $(x_1, x_2, \ldots, x_n)$ may take.

Ex:

$$ T_1(x_1, x_2, \ldots, x_n) = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \quad ; \quad T_2(x_1, x_2, \ldots, x_n) = x_{(1)} $$

$$ T_2(x_1, x_2, \ldots, x_n) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \quad ; \quad T_4(x_1, x_2, \ldots, x_n) = x_{(n)} $$

The only restriction on $T$, is that it should not depend on $\theta$. Ex: $\frac{x - \mu}{\sigma/\sqrt{n}}$ is not a statistic.
As a function of \( X_1, X_2, \ldots, X_n \), \( T(X_1, X_2, \ldots, X_n) \) is a r.v.
the properties of \( T \) will depend on the "parental" dist \( f(x) \) (run through trace of r.v.s).

**Ex. (Thm 5.2.6).** If \( X_1, X_2, \ldots, X_n \) is a random sample where \( E(X) = \mu \) and \( \text{Var}(X) = \sigma^2 \) then
\[
E(X) = \mu, \quad \text{Var}(X) = \sigma^2/n, \quad E(S^2) = \sigma^2 \quad \text{where} \quad S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2
\]
(\( \bar{X} \) and \( S^2 \) are unbiased to estimate \( \mu \) and \( \sigma^2 \) respectively)

**Ex. (Thm 5.2.7)** If \( X_1, X_2, \ldots, X_n \) is a random sample with mgf \( M_X(t) = E(e^{tx}) \)
\[
M_X(t) = \left[ M_X(t/n) \right]^n
\]
very helpful to derive the dist. of \( \bar{X} \) easily

If \( X \sim N(\mu, \sigma^2) \) \( \Rightarrow M_X(t) = \exp(\mu t + \frac{\sigma^2 t^2}{2}) \)
\[
M_X(t) = \left[ \exp(\mu t/n + \frac{\sigma^2 (t/n)^2}{2}) \right]^n
\]

\[= \exp(\mu t + \frac{\sigma^2 (t/n)^2}{2})\]
\[\Rightarrow \bar{X} \sim N(\mu/n, \sigma^2/n).
\]

If \( X \sim \text{Exp}(\lambda) = f(x; \lambda) = \frac{1}{\lambda} e^{-x/\lambda} \); \( \chi > 0 \)
\[
M_X(t) = \frac{1}{1 - \lambda t}
\]
\[\Rightarrow M_X(t) = \left( \frac{1}{1 - \lambda(t/n)} \right)^n = \text{Gamma dist. with pa.rms } \alpha = n \beta = \frac{n}{\lambda}.
\]

\( \text{(Gamma}(\alpha, \beta) = f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha) \beta^\alpha} x^{\alpha-1} e^{-x/\beta}) \)

What if the mgf does not exist?

With iid \( f_X(x) = n f_{X_1, X_2, \ldots, X_n}(x) \) (see Ex 5.5).
OLD WINE STATISTICS
Given $X_1, X_2, \ldots, X_n$ a random sample; place values in ascending order
$X(1), X(2), \ldots, X(n)$
$X(i) = \min \{X_i\} \ ; \ X(i) = \text{second smallest } X_i \ ; \ X(n) = \max \{X_i\}$

SAME RANGE: $R = X(n) - X(1)$
SAME MEDIAN: $M(n)$
$X(n), X(n), X(3) \rightarrow M = X(2)$
$X(n), X(n), X(n), X(4) \rightarrow M = \frac{X(3) + X(4)}{2}$

In general
$$M = \begin{cases} X(n + \frac{1}{2}) & \text{if } n \text{ is odd} \\ \frac{X(n + \frac{1}{2}) + X(n + \frac{1}{2} + 1)}{2} & \text{if } n \text{ is even} \end{cases}$$

$M$ is a measure of location, less sensitive than the mean to extreme observations.

MAIN RESULT (Theorem 5.4.1)

If $X_1, X_2, \ldots, X_n$ denote the old wine statistic of a random sample
$X_1, X_2, \ldots, X_n$ of continuous variables with cdf $F_X(x)$ and pdf $f_X(x)$, then, the pdf of $X(j)$ is

$$f_X(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) [F_X(x)]^{j-1} [1 - F_X(x)]^{n-j}$$

$X(j)$ is $x$ if at least $j$ $X_i$'s are less than or equal to $x$

Let $Y = \text{no. of } X_i$'s $\leq x$.

$$X(j) \leq x \iff \{X \leq x \} \text{ is the event} \iff Y \leq j \text{ and } Y \sim \text{Bin}(n, F_X(x))$$

$$P[X(j) \leq x] = \sum_{k=j}^{n} \binom{n}{k} [F_X(x)]^k [1 - F_X(x)]^{n-k}, \quad k = 0, 1, \ldots, n$$
\[ f(x_0) = \frac{d}{dx} P[X_0 < x_0] \]

\[ = \sum_{k=0}^{n} \binom{n}{k} F(X_0)^{k-1}(1-F(X_0))^{n-k} f(x_0) \]

\[ = \sum_{k=j}^{n} \binom{n}{k} \int_{x_0}^{x_0} F(X_0)^{k-1}(1-F(X_0))^{n-k} f(x_0) \]

\[ = \binom{n}{j} \int_{x_0}^{x_0} F(X_0)^{j-1}(1-F(X_0))^{n-j} f(x_0) + \sum_{k=j+1}^{n} \binom{n}{k} \int_{x_0}^{x_0} F(X_0)^{k-1}(1-F(X_0))^{n-k} f(x_0) \]

\[ \times \binom{k}{j} \sum_{k=0}^{j} \frac{F(X_0)^{k}}{k!} \]

Note that \( \binom{n}{j} j = \frac{n!}{(n-j)! j!} j = \frac{n!}{(n-j)! (j-1)!} \)

Change of Variables \( k = j + 1 \), \( n \)

\[ k' = k - 1 \quad \text{so} \quad k' = j, j + 1, \ldots, n - 1 \]

\[ \Theta = \sum_{k'=j}^{n} \binom{n}{k'+1} (k'+1) \int_{x_0}^{x_0} F(X_0)^{k'}(1-F(X_0))^{n-k'-1} f(x_0) \]

Since \( k' \) is a dummy variable we may replace \( k' \) with \( k \).

Just wish to check that

\[ \binom{n}{k} (n-k) = \binom{n}{k+1} (k+1) \]

\[ \frac{n!}{k! (n-k)!} (n-k) = \frac{n!}{k! (n-k-1)!} \quad \text{so} \quad \Theta = 0 \]

Particular Cases

\( j = 1 \), \( j = n \)

\[ f(x_0) = \frac{n!}{(n-1)!} f(x_0) [1-F(x_0)]^{n-2} = n f(x_0) [1-F(x_0)]^{n-1} \]

\[ f(x_0) = \frac{n!}{(n-1)!} f(x_0) [F(x_0)]^{n-2} [1-F(x_0)]^{n-1} = n f(x_0) [F(x_0)]^{n-1} \]

\( x_1, x_2, \ldots, x_n \) are i.i.d. from \( U(0,1) \)

\[ f(x) = 1 \quad 0 < x < 1 \]

\[ f_{x|y} (x) = n (1-x)^{n-1} \]

\[ f_X(x) = x \]

\[ f_{x|y} (x) = n x^{n-1} \]
For auxiliary $j$

\[ f_{X(n), Y(n)}(u,v) = \binom{n-1}{j-1} \frac{(n-j)!}{(n-j)!} \frac{u^{j-1}}{(1-u)^{n-j+1}} \]

\[ E(\mathbf{Z}) = \frac{\alpha \beta}{(d+\beta)^2} \quad \text{Var}(\mathbf{Z}) = \frac{\alpha \beta}{(d+\beta+\alpha)^2} \]

Therefore

\[ E(X(n)) = \frac{(\alpha \beta)}{(n+\alpha)} \quad \text{Var}(X(n)) = \frac{(\alpha \beta)(n-\beta+1)}{(n+\alpha)^2(n+\beta)} \]

Theorem 6.4.6

Gives an expression for

\[ f_{X(n), X(j)}(u,v) \quad 1 \leq j \leq n \]

But we know $v = j$ for $j \leq n$ Eq 5.4.7 gives

\[ f_{X(n), X(j)}(u,v) = \binom{n-1}{j-1} \frac{u^{j-1}}{(1-u)^{n-j+1}} \frac{v^{n-j}}{(1-v)^j} \]

\[ E(X(n)) = \frac{(\alpha \beta)}{(n+\alpha)} \quad \text{Var}(X(n)) = \frac{(\alpha \beta)(n-\beta+1)}{(n+\alpha)^2(n+\beta)} \]

Theorem 6.4.6

Gives an expression for

\[ f_{X(n), X(j)}(u,v) \quad 1 \leq j \leq n \]

But we know $v = j$ for $j \leq n$ Eq 5.4.7 gives

\[ f_{X(n), X(j)}(u,v) = n(n-1)(\frac{1}{\alpha}) (\frac{1}{\alpha}) \left[ \frac{v}{\alpha} - \frac{u}{\alpha} \right]^{n-2} \]

\[ = n(n-1)(v-u)^{n-2} \quad 0 < u < v < \alpha \]

Also

\[ f_{X(n)}(v) = n(\frac{1}{\alpha}) (\frac{v}{\alpha})^{n-1} = \frac{v^{n-1}}{\alpha^n} \]

Theorem 6.4.6

\[ f_{X(n), X(j)}(u,v) = \frac{f_{X(n), X(j)}(u,v)}{f_{X(n)}(v)} = \frac{(n-1)(v-u)^{n-2}}{v^{n-1}} \]

Endpoints of $X(n)$ given $X(n)$

Ex 5.27.
Converging Concepts.

Why? Long-run behavior (n→∞) of sample quantities such as \( \bar{x} \) or \( s^2 \) relates to asymptotics. Permits approximations to "large" n (but fixed) situations.

3 Types of Convergence: i) Convergence in Probability ii) Almost Sure Convergence iii) Convergence in Distribution

**Def. 1: A sequence of random variables \( X_1, X_2, \ldots \) converges in probability to \( X \) if for every \( \epsilon > 0 \)

\[
\lim_{n \to \infty} P(|X_n - X| > \epsilon) = 0 \quad \text{or} \quad \lim_{n \to \infty} P(|X_n - X| < \epsilon) = 1
\]

We denote this by \( X_n \xrightarrow{P} X \).

Ex: \( X_n \sim N(0, 1/n) \) \( \Rightarrow \) \( X_n \xrightarrow{P} 0 \)

Why?

Consider \( P(|X_n| < \epsilon) = P(-\epsilon < X_n < \epsilon) \)

\[
= P\left(-\epsilon/\sqrt{1/n} < Z < \epsilon/\sqrt{1/n}\right) \quad \text{(1) where } Z \sim N(0, 1)
\]

As \( n \to \infty \)

(1) = \( P(-\infty < Z < \infty) = 1 \)

Ex. 2 (Weak Law of Large Numbers). Let \( X_1, X_2, \ldots, X_n \) iid random variables with \( E(X_i) = \mu \) and \( \text{Var}(X_i) = \sigma^2 \), if \( \bar{X}_n = \frac{1}{n} \sum X_i \) then, for every \( \epsilon > 0 \)

\[
\lim_{n \to \infty} P(|\bar{X}_n - \mu| < \epsilon) = 1
\]

Proof (via Chebyshev's Ineq) For every \( \epsilon > 0 \)
\[ P(1X_n - \mu < \varepsilon) \rightarrow P((\bar{X}_n - \mu)^2 > \varepsilon^2) \leq \frac{\text{Var}(\bar{X}_n - \mu)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \]

\[ \Rightarrow P(1X_n - \mu < \varepsilon) \rightarrow 1 - \frac{\sigma^2}{n\varepsilon^2} \quad \text{as} \quad n \rightarrow \infty \quad \text{Proof} \rightarrow \text{1} \]

For \( s^2 \):
\[ s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \quad \text{Can we prove a WN?} \]

\[ P(1Sn^2 - s^2 < \varepsilon) \leq \frac{\text{Var}(Sn^2)}{\varepsilon^2} = \frac{\text{Var}(s^2)}{\varepsilon^2} \quad \text{as long as Var}(s^2) \rightarrow 0 \]

Almost sure convergence:

Def 2: A sequence of random variables \( X_1, X_2, \ldots \)

\( \text{Converges} \quad \text{as} \quad \text{almost surely} \quad \text{to} \quad X, \quad \text{if every} \quad \varepsilon > 0 \)

\[ P(\lim_{n \rightarrow \infty} |X_n - X| < \varepsilon) = 1 \]

We denote this by \( X_n \rightarrow^a X \)

Result: Almost sure convergence implies convergence in probability, but not the opposite.

Ex: 5.5.8

**Consider a sample space** \( S = [0, 1] \) with a \( U(0,1) \) distribution.
For any point \( s \in S \), we define
\[
X_1(s) = s + 1_{[0,1]}(s); \quad X_2(s) = s + 1_{[0,1/2]},
\]
\[
X_3(s) = s + 1_{[1/2,1]}(s); \quad X_4(s) = s + 1_{[0,1/3]},
\]
\[
X_5(s) = s + 1_{[1/3,2/3]}(s); \quad X_6(s) = s + 1_{[2/3,1]},
\]
\[
X_7(s) = s + 1_{[0,1/4]}(s); \quad X_8(s) = s + 1_{[1/4,1/2]}.
\]

Let's make \( X(s) = s \) (identity function).
\[
P\left(\left|X_n(s) - X(s)\right| > \varepsilon\right) = P\left(|1_n| = \text{length}(I_n)\right)
\]

But as \( n \to \infty \), \( \text{length}(I_n) \to 0 \) since \( X_n \to X \).
For a.s., we need \( X_n(s) \to s \) except on a set with measure zero.
Notice that for any \( s \), \( X_n(s) \to s + 1, \) or \( s \).
Take \( s = 1/2 \): \( X_1(1/2) = 3/2 \), \( X_2(1/2) = 3/2 \), \( X_3(1/2) = 3/2 \).
\( X_4(s) = 1/2 \), \( X_5(s) = 3/2 \).

Notice that the subsequence \( X_{n_k}(s) = s + 1_{[1/2,1]}(s) \) converges both in prob. and a.s.

For large enough \( n \), \( X_n(s) = s \).


Strong Law of Large Numbers. Let \( X_1, X_2, \ldots \) be i.i.d. RVs with mean \( \mu \) and variance \( \sigma^2 \). Then, for every \( \varepsilon > 0 \),
\[
P\left(\lim_{n \to \infty} \left|\frac{1}{n} \sum_{i=1}^{n} X_i - \mu\right| < \varepsilon\right) = 1 \quad (X_n \xrightarrow{a.s.} \mu)
\]
LAST FORM OF CONVERGENCE.

CONVERGENCE IN DISTRIBUTION.

Def 3. A sequence of random variables $X_1, X_2, \ldots$ converges in distribution to a random variable $X$ if
\[ \lim_{n \to \infty} F_{X_n}(x) = F_X(x) \]
ono
Denoted by $X_n \xrightarrow{D} X$.

Example: $Z \sim N(0, 1)$, $X_1 = 2, X_2 = -2, X_3 = 2$.

Notice that $-Z \sim N(0, 1)$ and then $X_n \xrightarrow{D} Z$.

Can this sequence converge in probability to $Z$?

Notice that:
\[ P(\lvert X_n - Z \rvert \leq \varepsilon) = \begin{cases} 1 & \text{if } X_n = Z \quad (n \text{ odd}) \\ \frac{1}{2} \left(1 \pm \frac{1}{2} \varepsilon\right) & \text{if } X_n = -Z \quad (n \text{ even}). \end{cases} \]

\[ \Rightarrow X_n \text{ does not converge in probability.} \]

Theo. If $X_1, X_2, \ldots$ $X_n$ converges in prob. to $X$, the sequence also converges in dist. to $X$.

Proof. (Ex. 5.40) Cont. Case. We can show that
\[ P(X \leq t - \varepsilon) \leq P(X_n \leq t) + P(\lvert X_n - X \rvert \geq \varepsilon) \quad (1) \]

since $X_n \leq t \iff \lvert X_n - X \rvert \geq \varepsilon$ implies $X - X_n \leq -\varepsilon + \varepsilon \leq t - \varepsilon$.

In the same manner:
\[ P(X_n \leq t) \leq P(X \leq t + \varepsilon) + P(\lvert X_n - X \rvert \geq \varepsilon) \quad (2) \]

Combining (1) and (2), we have
\[ P(X \leq t - \varepsilon) - P(\lvert X_n - X \rvert \geq \varepsilon) \leq P(X \leq t) \leq P(X \leq t + \varepsilon) + P(\lvert X_n - X \rvert \geq \varepsilon). \]

As $n \to \infty$, $P(X \leq t - \varepsilon) \leq P(X \leq t) \leq P(X \leq t + \varepsilon)$
\[ F_X(t - \varepsilon) \leq F_{X_n}(t) \leq F_X(t + \varepsilon), \]

\[ \Rightarrow \lim_{n \to \infty} F_{X_n}(x) = F_X(x). \]
Famous example of convergence in dist.

Central Limit Theorem: \( X_1, X_2, \ldots \) are iid Rvs.
with \( E(X_i) = \mu \) and \( Var(X_i) = \sigma^2 \).
If \( Z_n = \frac{X_n - \mu}{\sigma} \) and 
\( F_{Z_n}(z) \) is the dist. function of \( Z_n \) then
\[
\lim_{n \to \infty} F_{Z_n}(z) = \Phi(z)
\]
with
\[
\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx
\]

Book offers proof when \( \mu_X(t) \) in a neighborhood of 0.
But the result is more general.

Ex. (typical of 345)

Suppose \( X_1, X_2, \ldots, X_{10} \) are iid \( X_i \sim Poisson(\lambda) \).
\( f(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}; \lambda = 1, 2, \ldots \).
Suppose \( \lambda = 10 \).

\[
\sum_{i=1}^{10} X_i \sim Poisson(100)
\]

\( P(\sum_{i=1}^{10} X_i < 325) \)

By mgf \( M_{\sum X_i(t)} = E(e^{t \sum X_i}) = [M_{X_i(t)}]^n \)
\( = e^{t \lambda} \sim Poisson(\lambda) \)

\( M_{X_i(t)} = e^{t(e^{\lambda} - 1)} \Rightarrow [M_{X_i(t)}]^n = e^{n \lambda(e^{t} - 1)} \)

\( \sum_{i=1}^{10} X_i \sim Poisson(100) \) \( \lambda = 100 \)

Exact Calculation \( P(\sum_{i=1}^{10} X_i < 325) = 0.928 \pm 0.27 \)

\( P(\sum_{i=1}^{10} X_i < 325) = P(\bar{X} < 10.8333) \approx P(z < \frac{8.3333}{10.8333}) = 0.9255419 \)

One problem with CLT is that we really don't know
how large \( n \) to use for the approx. to work.

(Exponential of \( f(x) \)).
Suppose now that we care about $g(X)$, some function of $X$. Do we have some sort of CLT for $g(X)$?

In the Poisson example, it seems natural to estimate $\lambda$ (mean) with $\bar{X}$. But now if we want to estimate $e^{-\lambda} = P(X=0)$, an estimator would be $e^{-\bar{X}} = g(X)$.

Limit dist. for $e^{-\bar{X}}$?

The Delta Method:

Suppose we have a r.v. (statistic) $T$ such that $E(T) = \theta$ and $g'(\theta)$ exists and is different from zero at an observed value of $T$.

Let's take the Taylor series expansion of $g$ around $\theta$.

$g(t) = g(\theta) + g'(\theta) (t-\theta) + \text{Remainder}$

$\Rightarrow g(t) \approx g(\theta) + g'(\theta) (t-\theta)$

If we take expectations,

$E(g(T)) \approx g(\theta) + g'(\theta) E(T - \theta) = g(\theta)$

Similarly,

$E(g(T) - g(\theta))^2 \approx (g'(\theta))^2 E(T - \theta)^2$

$= \text{Var}(g(T)) \approx g'(\theta)^2 \text{Var}(T)$

We obtain approx. expressions for $E(g(T))$ and $\text{Var}(g(T))$.

In the ex., $\theta = \lambda$

$E(e^{-\bar{X}}) \approx e^{-\lambda}$ and $\text{Var}(e^{-\bar{X}}) \approx e^{-2\lambda} \left( \frac{\lambda}{n} \right)$

$g(T)$ is "approx. unbiased."

In fact, if $T \tilde{\rightarrow} T_n$ is a sequence of R.V. where

$\sqrt{n}(T_n - \theta) \tilde{\rightarrow} N(0, \sigma^2)$. If $g$ is a function whose $g'$ exists and is not 0 then:

$E(g(T)) \approx g(\theta)$ and $\text{Var}(g(T)) \approx g'(\theta)^2 \text{Var}(T)$.
\[
\sqrt{n} \left( g(\bar{X}) - g(\theta) \right) \xrightarrow{d} N(0, \sigma^2 g'(\theta)^2)
\]

In terms of the example:
\[
\sqrt{n} \left( e^{-\bar{X}} - e^{-\theta} \right) \xrightarrow{d} N(0, e^{-2\theta} \frac{\theta}{n})
\]

Usually \(\sigma^2\) and \(\theta\) are unknown, hence \(\sigma^2 g'(\theta)^2\) is unknown. If we can find \(V_n\) such that \(V_n \xrightarrow{d} \sigma^2 g'(\theta)^2\)

Then Slutsky's theorem (PAG. 240) guarantees that:
\[
\sqrt{n} \frac{g(\bar{X}) - g(\theta)}{\sqrt{\sigma^2 g'(\theta)^2}} \xrightarrow{d} N(0, 1)
\]

Does not involve \(\theta\) in the theorem.

In the example Poisson, by the WLLN \(\bar{X} \xrightarrow{P} \lambda\)

By M.W. ex. prob. \(\bar{X} = e^{-2\theta} \left( \frac{\bar{X}}{\theta} \right)^{\theta} \xrightarrow{P} e^{-2\theta} \left( \frac{\lambda}{\theta} \right)^{\theta}\)

Then
\[
\sqrt{n} \left( e^{-\bar{X}} - e^{-\theta} \right) / \sqrt{\sigma^2 g'(\theta)^2} \xrightarrow{d} N(0, 1)
\]

Justifies the use of \(\bar{X}\) plug in as estimator for \(\lambda\).

Finally, what if \(g'(\theta) = 0\)

second order exp. on \(g\)
\[
g(\bar{X}) = g(\theta) + g'(\theta) (\bar{X} - \theta) + \frac{1}{2} g''(\theta) (\bar{X} - \theta)^2 + R
\]

\(\Rightarrow g(\bar{X}) \approx \frac{1}{2} g''(\theta) (\bar{X} - \theta)^2 + g(\theta)
\]

\[n(g(\bar{X}) - g(\theta)) \times \frac{\bar{X} - \theta}{\sqrt{\sigma^2 g'(\theta)^2}} \xrightarrow{d} N(0, 0.2)
\]

\[n(g(\bar{X}) - g(\theta)) \rightarrow \sigma^2 g''(\theta) \times 2
\]

\[n(g(\bar{X}) - g(\theta)) \rightarrow \sigma^2 g''(\theta) \times 2
\]
Generating Random Samples.

In inferring, we are interested in the distribution of \( T(x_1, x_2, \ldots, x_n) \) a "statistic".

Suppose it is very hard to deal with the dist of \( T \) with pdf or conjugate concepts. (For ex. an order statistic).

Idea:
Since \( x_1, x_2, \ldots, x_n \) are iid where \( x_i \sim f(x_i) \)
we can "generate" with the help of the computer values for \( x_1, x_2, \ldots, x_n \). To obtain values for \( T(x_1, x_2, \ldots, x_n) \).

Ex:

Let \( x_1, x_2, \ldots, x_{20} \) be iid \( N(0, 1) \) RVs.

And \( T = \max \{x_1, x_{20}\} \)

Hard to deal with the dist of \( T \) because \( F_T(t) = P(X_{20} \leq t) = \prod_{i=1}^{20} P(x_i \leq t) = [\Phi(t)]^{20} \).

And \( \Phi(t) \) does not have a closed form expression.

Generate values.

1. \( - \quad x_1^{(1)}, x_2^{(1)}, x_{20}^{(1)} \quad \rightarrow \quad x_{(20)}^{(1)} \quad \text{and repeat and on.} \)

2. \( - \quad x_1^{(2)}, x_2^{(2)}, x_{20}^{(2)} \quad \rightarrow \quad x_{(20)}^{(2)} \)

\( \vdots \)

(1000), (1000), (1000), (1000)

1000, x_1, x_2, x_{20} \quad \rightarrow \quad x_{(20)}^{(1000)}

SUMMARIZE: \( x_{(20)}, x_{(20)}, \ldots, x_{(20)} \).

Look at handout. Fig (a)
In fact, if we want to approximate \( E(X(20)) \) and \( E(X^2(20)) \):

\[
E(X(20)) \approx \frac{1}{1000} \sum_{i=1}^{1000} X(i) = 1.867 \quad (\text{with my simulation})
\]

\[
E(X^2(20)) \approx \frac{1}{1000} \sum_{i=1}^{1000} (X(i))^2 = 3.786
\]

**Issue:** How do we generate \( X \) with a pdf \( f(x|10) \)?

Start with an algorithm that generates \( U \); \( U \) follows a \( U(0,1) \) distribution.

For \( X \) continuous, we generate \( X \) by

\[
X = F^{-1}(U)
\]

where \( F^{-1} \) is the inverse CDF of \( F \).

**Example:** If \( X \) is \( \text{Exp}(\lambda) \) then \( f(x|\lambda) = \frac{1}{\lambda} e^{-x/\lambda}, \ x > 0 \)

so

\[
F(x|\lambda) = 1 - e^{-x/\lambda}; \quad \text{if we take}
\]

\[
U = 1 - e^{-X/\lambda} \implies e^{-X/\lambda} = 1 - U
\]

\[
\Rightarrow \quad -\frac{X}{\lambda} = \log(1-U) \implies X = -\lambda \log(1-U)
\]

**Example:** \( F^{-1} \) may not have a closed form (box-muller).

**For \( X \) discrete:**

Suppose \( X \) takes 3 values: \( x_1 \) with prob. \( p_1 \), \( x_2 \) with prob. \( p_2 \) and \( x_3 \) with prob. \( p_3 \).

To generate a value of \( X \):

(i) Generate \( U \) from \( U(0,1) \)

(ii) If \( U \leq p_1 \) \( \implies X = x_1 \)

\[
\begin{align*}
\text{If } & \quad p_1 < U \leq p_1 + p_2 \quad \implies X = x_2, \\
& \quad \frac{F(x_1)}{F(x_2)}
\end{align*}
\]
Most statistical packages have a simulator for well-established \( f(x|\theta) \). (See Figs. (a)-(d)).

Still, if \( f(x|\theta) \) is not available, we may use indirect methods.

\( X \sim \text{DEXP}(\lambda = 0) \). So \( f(x|\mu) = \frac{1}{2} e^{-\frac{1}{2} (x-\mu)^2} \) \( -\infty < x < \infty \) \( (\mu = 1) \). Suppose we want to generate values of \( X \).

Proposed auxiliary density. (Candidate) 
\( f(y|\mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} (y-\mu)^2\right) \) \( -\infty < y < \infty \)

Generate \( Y \sim \text{N}(0,1) \), set \( X_0 = Y \)

For \( i = 1, 2, \ldots \)

1. Generate \( U_i \sim \text{U}(0,1) \); \( Y_i \sim \text{N}(0,1) \).

\[ P_i = \min\{1, \frac{\exp\left(-\frac{1}{2} (y_i)\right) \exp\left(-\frac{1}{2} (x_0)^2\right)}{\exp\left(-\frac{1}{2} y_i^2\right) \exp\left(-\frac{1}{2} (x_0)^2\right)}\} \]

2. Set 
\[ X_i = \begin{cases} Y_i & \text{if } U_i \leq P_i \\ X_{i-1} & \text{if } U_i > P_i \end{cases} \]

The key thing is that as \( i \to \infty \) \( X_i \to X \)

This leads to the Metropolis-Hastings algorithm.