Properties of a Random Sample

Stat 205B

Department of Statistics
University of California Santa Cruz

Winter, 2022

Notes

Recap: Expectation

For a continuous random variable,

$$E\{g(X)\} = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

Properties:

- $E(a) = a$ for any constant $a \in \mathbb{R}$
- $E\{ag(X)\} = aE\{g(X)\}$ for any constant $a \in \mathbb{R}$
- $E\{h(X) + g(X)\} = E\{h(X)\} + E\{g(X)\}$

Notes

Recap: Variance

Notes

The variance of a random variable is

$$\mathrm{Var}(X) = \mathrm{E}\left[\{X - \mathrm{E}(X)\}^2\right] = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2$$

Properties:

- $\mathrm{Var}(a) = 0$ for any constant $a \in \mathbb{R}$
- $\mathrm{Var}\{ag(X)\} = a^2\mathrm{Var}\{g(X)\}$ for any constant $a \in \mathbb{R}$

Recap: Covariance

Notes

The covariance of two random variables, X and Y, is

$$\mathrm{Cov}(X, Y) = \mathrm{E}\left[\{X - \mathrm{E}(X)\}\{Y - \mathrm{E}(Y)\}\right] = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$$

Properties:

- $\mathrm{E}\{h(X,Y) + g(X,Y)\} = \mathrm{E}\{h(X,Y)\} + \mathrm{E}\{g(X,Y)\}$
- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$
- If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$ and $\mathrm{E}\{h(X)g(Y)\} = \mathrm{E}\{h(X)\}\,\mathrm{E}\{g(Y)\}$

Random Sample

Definition:
Random variables $X_1, \ldots, X_n$ are said to be a random sample of size n from $f(x)$ if $X_1, \ldots, X_n$ are mutually independent and the marginal pdf or pmf of each X_i is $f(x)$.

We also use the terminology $X_1, \ldots, X_n$ are "independent and identically distributed (iid)" with pdf or pmf $f(x)$.

Notes

Random Sample Examples

Two common situations in which a random sample is an appropriate model:

- $X_1, \ldots, X_n$ are the measurements resulting from n identical experiments that are conducted independently of each other. Here, X_i may be either discrete or continuous.
- Take a simple random sample of size n with replacement from a population having a finite number of elements. X_i is the number selected on the ith draw. Here, X_i is discrete.

Notes

Populations

When $X_1, \ldots, X_n$ area a random sample from a pdf or pmf $f(x)$, then $f(x)$ is often referred to as a population.

The joint pdf or pmf of $X_1, \ldots, X_n$ is:

$$\begin{aligned} f(x_1, \ldots, x_n) &= f(x_1)f(x_2)\cdots f(x_n) \\ &= \prod_{i=1}^{n} f(x_i) \end{aligned} \tag{1}$$

If $f(x)$ is a member of a parametric family, we can write

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta})$$

Notes

Typical Statistical Setting

We know that a certain population is in a parametric family of distributions, but we do not know the actual parameter value. We perform a sequence of experiments or take a SRS with replacement, yielding $X_1, \ldots, X_n$, a random sample from $f(\cdot|\boldsymbol{\theta})$. We then use the random sample to make an educated guess as to the value of $\boldsymbol{\theta}$.

$X_1, \ldots, X_n$ will "behave" differently depending on the underlying value of $\boldsymbol{\theta}$. This forms the basis for making an educated guess.

Notes

Example 1

Let $X_1, \ldots, X_n$ be a random sample from a Unif$(0, \theta)$ distribution, where $\theta > 0$. Define

$$Y = \max(X_1, \ldots, X_n). \tag{2}$$

Clearly, Y will be closer to θ than any other value of X_i will be. To find out how close Y tends to be to θ, we can determine the probability distribution of Y.

Notes

Example 1 Cont.

Let $0 < y < \theta$. Then,

$$\begin{aligned} P(Y \le y) &= P(X_1 \le y, X_2 \le y, \ldots, X_n \le y) \\ &= \prod_{i=1}^{n} P(X_i \le y) \\ &= \left(\frac{y}{\theta}\right)^n \end{aligned}$$

$$F_Y(y) = \begin{cases} 0, & y \le 0 \\ \left(\frac{y}{\theta}\right)^n, & 0 < y \le \theta \\ 1, & y > \theta. \end{cases}$$

Notes

Example 1 Cont.

How likely is it that Y will be within $\epsilon\theta$ of θ $(0 < \epsilon < 1)$?

$$\begin{aligned} P\left(\theta(1-\epsilon) < Y < \theta\right) &= P(Y < \theta) - P(Y < \theta(1-\epsilon)) \\ &= 1 - (1-\epsilon)^n \end{aligned}$$

Note that this tends to 1 as $n \to \infty$.

Suppose $\epsilon = 0.1$ and $n = 10$.

$$P(0.9\theta < X_i < \theta) = 0.1$$

$$P(0.9\theta < Y < \theta) = 1 - (0.9)^{10} \approx 0.65$$

Notes

Statistics

Definition:
Let $X_1, \ldots, X_n$ be a random sample from a population and let $T(x_1, \ldots, x_n)$ be a real-valued or vector-valued function that does not depend on unknown parameters, and whose domain contains the range of $(X_1, \ldots, X_n)$. Then the random vector $Y = T(X_1, \ldots, X_n)$ is called a statistic, and the probability distribution of Y is called the sampling distribution of Y.

Notes

Example 2

$Y = \max(X_1, \ldots, X_n)$ in (2) is a statistic.

Note: The function $\max(x_1, \ldots, x_n)$ does not depend on any unknown parameters. Hence, we can compute the value of Y as soon as we know $x_1, \ldots, x_n$, the realized sample values.

The probability distribution derived in Example 1 is the sampling distribution of the statistic Y.

Notes

More Statistics

Other examples of statistics:

- Sample mean: $\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$
- Sample variance: $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$
- kth sample moment: $\sum_{i=1}^{n} \frac{X_i^k}{n}$

Notes

Sums of iid Random Variables

Important properties of sums of iid random variables:
Let g be a function such that $\mathrm{Var}\{g(X_i)\}$ exists. Then

$$\mathrm{E}\left\{\sum_{i=1}^{n} g(X_i)\right\} = n\mathrm{E}\{g(X_1)\} \tag{3}$$

$$\mathrm{Var}\left\{\sum_{i=1}^{n} g(X_i)\right\} = n\mathrm{Var}\{g(X_1)\} \tag{4}$$

Notes

Proof

$X_1, \ldots, X_n$ are mutually independent, so

$$\mathrm{Var}\left\{\sum_{i=1}^{n} g(X_i)\right\} = \sum_{i=1}^{n} \mathrm{Var}\{g(X_i)\}.$$

Then, since X_i's are identically distributed,

$$\mathrm{Var}\{g(X_i)\} = \mathrm{Var}\{g(X_1)\}, \; i = 1, \ldots, n$$

and hence

$$\mathrm{Var}\left\{\sum_{i=1}^{n} g(X_i)\right\} = n\mathrm{Var}\{g(X_1)\}.$$

Notes

Properties of $\bar{X}$ and S^2 in a Random Sample

Let $X_1, \ldots, X_n$ be. random sample from a population with mean μ and variance $\sigma^2 < \infty$. Then

$$E(\bar{X}) = \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) = \frac{1}{n} \cdot nE(X_1) = \mu.$$

$$\begin{aligned} Var(\bar{X}) &= \frac{1}{n^2} Var\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \cdot nVar(X_1) \\ &= \frac{\sigma^2}{n} \end{aligned}$$

Notes

Properties Cont.

$$\begin{aligned} E(S^2) &= \frac{1}{n-1} \sum_{i=1}^n E\{(X_i - \bar{X})^2\} \\ (X_i - \bar{X})^2 &= X_i^2 - 2X_i\bar{X} + \bar{X}^2 \\ E(X_i^2) &= Var(X_i) + \mu^2 = \sigma^2 + \mu^2 \\ E(\bar{X}^2) &= Var(\bar{X}) + \mu^2 = \sigma^2/n + \mu^2 \\ E(X_i\bar{X}) &= \frac{1}{n} E\left(X_i \sum_{j=1}^n X_j\right) = \frac{1}{n} \sum_{j=1}^n E(X_iX_j) \\ &= \frac{1}{n}\left\{E(X_i^2) + \sum_{j\neq i} E(X_iX_j)\right\} = \frac{1}{n}\{\sigma^2 + \mu^2 + (n-1)\mu^2\} \\ &= \frac{\sigma^2}{n} + \mu^2 \end{aligned}$$

Notes

Properties Cont.

Therefore,

$$\mathrm{E}\{(X_i - \bar{X})^2\} = \sigma^2 + \mu^2 - 2\left(\frac{\sigma^2}{n} + \mu^2\right) + \frac{\sigma^2}{n} + \mu^2$$
$$= \left(\frac{n-1}{n}\right)\sigma^2.$$

Finally[1],

$$\mathrm{E}(S^2) = \sigma^2$$

Since $\mathrm{E}(\bar{X}) = \mu$ and $\mathrm{E}(S^2) = \sigma^2$, $\bar{X}$ and S^2 are said to be unbiased estimators of μ and σ^2 respectively.

[1] The proof on pg. 214 (CB) uses the fact that $S^2 = \frac{1}{n-1}\left\{\sum_{i=1}^n X_i^2 - n\bar{X}^2\right\}$.

Notes

Sampling Distribution of $\bar{X}$

Let $X_1, \ldots, X_n$ be a random sample from a population having mgf M_X. Then we compute the mgf of $\bar{X}$ as

$$M_{\bar{X}}(t) = \mathrm{E}(e^{t\bar{X}})$$
$$= \mathrm{E}\left\{\exp\left(\frac{t}{n}\sum_{i=1}^n X_i\right)\right\} = \mathrm{E}\left\{\prod_{i=1}^n \exp\left(\frac{t}{n}X_i\right)\right\}$$
$$= \prod_{i=1}^n \mathrm{E}\left\{\exp\left(\frac{t}{n}X_i\right)\right\} = \prod_{i=1}^n M_X(t/n) = \{M_X(t/n)\}^n$$

If $\{M_X(t/n)\}^n$ is a "familiar" mgf, then we have found the probability distribution of $\bar{X}$.

Notes

MGF Examples

- $X_1, \ldots, X_n$ a random sample from $\mathrm{N}(\mu, \sigma^2)$, then $\bar{X} \sim \mathrm{N}(\mu, \sigma^2/n)$
- $X_1, \ldots, X_n$ a random sample from $\mathrm{gamma}(\alpha, \beta)$, then $barX \sim \mathrm{gamma}(n\alpha, \beta/n)$
- $X_1, \ldots, X_n$ a random sample from $\mathrm{Poisson}(\lambda)$, then

$$M_X(t) = \exp\left\{\lambda(e^t - 1)\right\}$$
$$M_{\bar{X}}(t) = \exp\left\{n\lambda(e^{t/n} - 1)\right\}$$

Then $\bar{X}$ has the same distribution as Y/n where $Y \sim \mathrm{Poisson}(n\lambda)$.

Notes

Convolutions

Let X and Y be independent and absolutely continuous random variables with densities f_X and f_Y respectively. The pdf of $Z = X + Y$ is

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy$$
$$= \int_{-\infty}^{\infty} f_Y(z-x) f_X(x) dx$$

The integral $\int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy$ is called the convolution of f_X and f_Y.

Notes

Example 3

Convolution of iid uniform random variables.
Let X and Y be iid Unif$(0,1)$. Find the pdf of $Z = X + Y$.

$$f_Z(z) = \int_{-\infty}^{\infty} I_{(0,1)}(z-y) I_{(0,1)}(y) dy$$
$$= \int_0^1 I_{(0,1)}(z-y) dy$$

If $z < 0$ or $z > 2$, $f_Z(z) = 0$.
Let $z \in (0,1)$, then $f_Z(z) = \int_0^z dy = z$.
Let $z \in (1,2)$, then $f_Z(z) = \int_{z-1}^1 dy = 1 - (z-1) = 2 - z$.
Thus,

$$f_Z(z) = (1 - |1-z|) I_{(0,2)}(z).$$

Notes

Example 3 Cont.

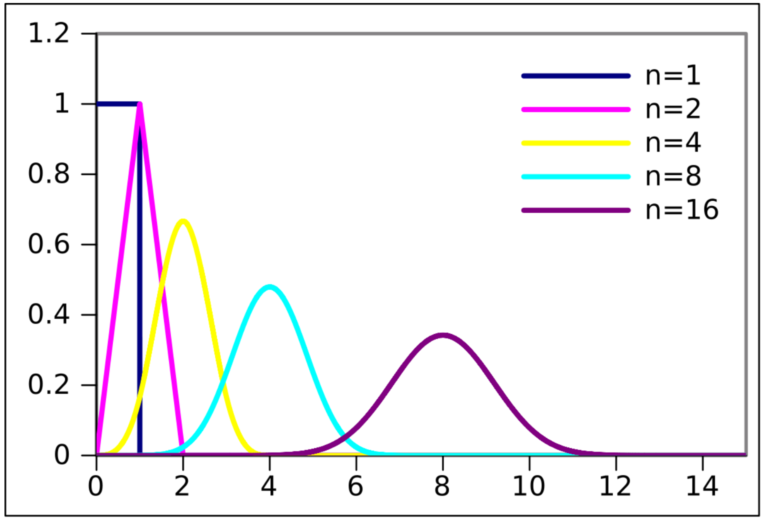


Figure: Irwin-Hall Distribution

Notes

Convolutions Cont.

$X_1 + X_n + \cdots + X_n$ is called an n-fold convolution when $X_1, \ldots, X_n$ are independent random variables.

Let $X_1, \ldots, X_n$ be a random sample. We can use the convolution formula to get the density of $X_1 + X_2$. Then, $(X_1 + X_2) + (X_3 + X_4)$ is the convolution of $X_1 + X_2$ with itself. We can use the convolution formula repeatedly to find the distribution of $X_1 + X_n + \cdots + X_n$ and hence $\bar{X}$.

The Central Limit Theorem says that the sum of n iid random variables is approximately normally distributed when n is large.

Notes

Convergence of a Sequence of Random Variables

To approximate various aspects of the sampling distributions of statistics, a common practice is to consider these aspects as n, the sample size, tends to ∞.

For example, suppose $X_1, \ldots, X_n$ is a random sample from a population with finite variance, σ^2 and population mean, μ. Then

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

and $\mathrm{E}(\bar{X}_n) = \mu$ and $\mathrm{Var}(\bar{X}_n) = \sigma^2/n$.

Does $\bar{X}_n$ converge (in some sense) to μ as $n \to \infty$? To answer this, we first need to define some convergence concepts.

Notes

Convergence in Probability

Definition:
Let $\{X_n : n = 1, 2, \ldots\}$ be any sequence of random variables. We say that $\{X_n\}$ converges in probability to a random variable X, if for each $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$$

It is equivalent to show that

$$\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1,$$

for each $\epsilon > 0$.

Does $\bar{X}_n$ converge to μ in probability?

Notes

Weak Law of Large Numbers

Suppose $X_1, \ldots, X_n$ are iid random variables with $\mathrm{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$, for $i = 1, 2, \ldots$
Define for each $n = 1, 2, \ldots,$

$$\bar{X}_n = \sum_{i=1}^{n} \frac{X_i}{n}.$$

Then $\bar{X}_n$ converges in probability to μ as $n \to \infty$.

Notes

Proof

We can use Markov's inequality.

$$P(|\bar{X}_n - \mu| \geq \epsilon) = P\left((\bar{X}_n - \mu)^2 \geq \epsilon^2\right)$$
$$\leq \frac{\mathrm{E}\left((\bar{X}_n - \mu)^2\right)}{\epsilon^2}$$
$$= \frac{\sigma^2}{n\epsilon^2}$$

Since this tends to 0 for each $\epsilon > 0$, $\bar{X}_n$ converges in probability to μ.

Notes

Almost Sure Convergence

Definition:
A sequence $\{X_n : \ n = 1, 2, \dots\}$ of random variables is said to converge almost surely to a random variable X if

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1.$$

Notes

Almost Sure Convergence Cont.

What does this mean? In the definition, it is implicit that $X, X_1, X_2, \ldots$ are all defined on the same sample space $\mathcal{S}$. Say we do the experiment and the result is, say, $s \in \mathcal{S}$. This yields a sequence of measurements, $X_1(s), X_2(s), \ldots$ and $X(s)$.

Then, $\{X_n(s) : \; n = 1, 2, \ldots\}$ is a sequence of real numbers that either converges to the real number $X(s)$ or it doesn't. We can divide $\mathcal{S}$ into two disjoint sets, $\mathcal{S}_c$ and $\mathcal{S} - \mathcal{S}_c$, where

$$\mathcal{S}_c = \left\{ s \in \mathcal{S} : \lim_{n \to \infty} X_n(s) = X(s) \right\}.$$

If $P(\mathcal{S}_c) = 1$ then we say that $X_n \to X$ almost surely (a.s.). Convergence with probability 1 is another term for almost sure convergence.

Notes

Convergence Relationship

Important Fact:
If X_n converges almost surely to X, then X_n converges in probability to X.

The converse is not necessarily true. See CB 5.5.8 for example.

Notes

Strong Law of Large Numbers

Let $X_1, X_2, \ldots$ be iid random variables with $\text{E}(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2 < \infty$.

Then $\bar{X}_n$ converges almost surely to μ as $n \to \infty$.

Notes

Convergence in Distribution

Definition:
$X_1, X_2, \ldots$ converge in distribution to a random variable X if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

for every point x at which F_X is continuous.

Convergence in distribution is very nearly point-wise convergence of the distribution function to F_X, the only exception being cases where F_X is not continuous at all points.

Notes

Convergence Relationship

Important Fact:
If the sequence of random variables $X_1, X_2, \ldots$ converges in probability to a random variable X, the sequence also converges in distribution to X.

Notes

Central Limit Theorem (CLT)

Let $X_1, X_2, \ldots$ be iid random variables with $\mathrm{E}(X_i) = \mu$ and $0 < \sigma^2 < \infty$. Define

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma},$$

and let $Z \sim \mathrm{N}(0, 1)$. Then, Z_n converges in distribution to Z.

Thus, regardless of the population we are sampling from, so long as the population has finite variance, $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ will have approximately a standard normal distribution for all n sufficiently large.

Notes

Slutsky's Theorem

If X_n converges to X in distribution and Y_n converges in probability to a constant a, then

1. $Y_n X_n$ converges to aX in distribution and
2. $X_n + Y_n$ converges to $X + a$ in distribution

Notes

Important Application of Slutsky's Theorem

Under the conditions of the CLT,

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to \mathrm{N}(0,1)$$

in distribution as $n \to \infty$.
In random samples where each X_i has finite fourth moment, one can show that S_n^2 converges in probability to σ^2. This implies that S_n converges in probability to σ. Then

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \frac{\sigma}{S_n}$$

Since $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ converges in distribution to $\mathrm{N}(0,1)$ and σ / S_n converges in probability to one, it follows from Slutky's theorem that

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \to \mathrm{N}(0,1)$$

in distribution.

Notes

Sampling from the Normal Distribution

Let $X_1, X_2, \ldots, X_n$ be a random sample from the $\mathrm{N}(\mu, \sigma^2)$ distribution. then we have the following properties:

- $\bar{X}_n \sim \mathrm{N}(\mu, \sigma^2/n)$
- $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$
- $\bar{X}$ and S^2 are independent random variables

Notes

Student's t Distribution

Again, let $X_1, \ldots, X_n$ be a random sample from the $\mathrm{N}(\mu, \sigma^2)$ distribution. the random variable

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

is important in the problem of statistical inference. It is used to infer (guess) the values of μ based on the information in $X_1, \ldots, X_n$.

What is the probability distribution of T?

Notes

Student's t Distribution Cont.

T is a function of $\bar{X}$ and $(n-1)S^2/\sigma^2$, whose joint distribution is

$$f(w,z) = \frac{1}{\sqrt{2\pi}\sigma/\sqrt{n}}\exp\left\{-\frac{1}{2}\left(\frac{w-\mu}{\sigma/\sqrt{n}}\right)^2\right\}\frac{z^{(n-1)/2-1}\exp\left(-\frac{z}{2}\right)}{\Gamma\left\{(n-1)/2\right\}2^{(n-1)/2}}I_{(0,\infty)}(z),$$

where $W = \bar{X}$ and $Z = (n-1)S^2/\sigma^2$. Then

$$T = \frac{W-\mu}{\sigma/\sqrt{n}}\sqrt{\frac{n-1}{Z}}.$$

Notes

Student's t Distribution Cont.

We may then find the density of T as follows:

1. Using the transformation method, find the density of (T, U), where $U = Z$
2. Having obtained the joint density $f_{T,U}$, find the marginal distribution of T in the usual way

$$f_T(t) = \int_0^\infty f_{T,U}(t,u)du$$

Notes

Student's t Distribution Cont.

The density of T is the so-called "t-distribution with $n-1$ degrees of freedom."

The Student's t density with k degrees of freedom is

$$f(t|k) = \frac{\Gamma\left\{\frac{k+1}{2}\right\}}{\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}} \left(\frac{1}{1+\frac{t^2}{k}}\right)^{\frac{k+1}{2}}$$

where $k = 1, 2, \ldots$

Notes

Order Statistics

The order statistics for a random sample $X_1, \ldots, X_n$ are simply $X_1, \ldots, X_n$ arranged in order from smallest to largest. These are denoted $X_{(1)}, \ldots, X_{(n)}$ where

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}.$$

The sample median, M, is defined in terms of order statistics:

$$M = \begin{cases} X_{\left(\frac{n+1}{2}\right)}, & n \text{ is odd} \\ \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2}, & n \text{ is even.} \end{cases}$$

The statistic $\max(X_1, \ldots, X_n)$ in Example 2 is one example of an order statistic, where $\max(X_1, \ldots, X_n) = X_{(n)}$.

Notes

Order Statistics Cont.

What if we want to find the cdf of $X_{(j)}$?

Let Y_x be the number of values in the random sample that are less than or equal to x. Then

$$P(X_{(j)} \leq x) = P(Y_x \geq j).$$

Now, $Y_x \sim \text{Bin}(n, F_X(x))$, where F_X is the cdf of each X_i. Hence,

$$P(X_{(j)} \leq x) = \sum_{k=j}^{n} \binom{n}{k} F_X(x)^k \left(1 - F_X(x)\right)^{n-k}.$$

Notes

Order Statistics Cont.

Now suppose the distribution of each X_i is absolutely continuous. then the distribution of $X_{(j)}$ is absolutely continuous with density

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) F_X(x)^{j-1} \left(1 - F_X(x)\right)^{n-j}$$

where f_X is the density of each X_i. See CB for the proof.

Notes

The Delta Method

Definition:
Suppose $g(x)$ has derivatives of order r, i.e. $g^{(r)}(x) = \frac{d^r}{dx^r} g(x)$ exists. Then for any constant a, the Taylor polynomial of order r about a is

$$T_r(x) = \sum_{i=0}^{r} \frac{g^{(i)}(a)}{i!}(x-a)^i.$$

Notes

The Delta Method Cont.

Theorem:
If $g^{(r)}(x) = \frac{d^r}{dx^r} g(x)|_{x=a}$ exists, then

$$\lim_{x \to a} \frac{g(x) - T_r(x)}{(x-a)^r} = 0$$

i.e. the approximation, $g(x) - T_r(x)$ always tends to zero faster than the highest order explicit term.

Since we are interested in approximations, we will ignore the remainder term.

Notes

The Delta Method Cont.

In statistics, we are usually concerned with the first order Taylor series (i.e. r=1). Let $T_1, \ldots, T_k$ be random variables with means $\theta_1, \ldots, \theta_k$ and

define $\boldsymbol{T} = (T_1, \ldots, T_k)$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$. Suppose there is a differentiable function $g(\boldsymbol{T})$ (an estimator of some parameter) for which we want an approximate estimate of the variance. Define

$$g_i'(\boldsymbol{\theta}) = \frac{\partial}{\partial t_i} g(\boldsymbol{t}) \Big|_{t_1=\theta_1,\ldots,t_k=\theta_k}$$

Notes

The Delta Method Cont.

The first order Taylor series expansion of g about $\boldsymbol{\theta}$ is

$$g(\boldsymbol{t}) = g(\boldsymbol{\theta}) + \sum_{i=1}^{k} g_i'(\boldsymbol{\theta})(t_i - \theta_i) + \text{Remainder}.$$

For our approximation, we omit the remainder and write

$$g(\boldsymbol{t}) \approx g(\boldsymbol{\theta}) + \sum_{i=1}^{k} g_i'(\boldsymbol{\theta})(t_i - \theta_i).$$

Notes

The Delta Method Cont.

Taking expectations,

$$\mathrm{E}_{\boldsymbol{T}} g(\boldsymbol{t}) \approx g(\boldsymbol{\theta}) + \sum_{i=1}^{k} g_i'(\boldsymbol{\theta}) \mathrm{E}_{\boldsymbol{T}}(t_i - \theta_i) = g(\boldsymbol{\theta}).$$

We can now approximate the variance of $g(\boldsymbol{T})$ by

$$\begin{aligned}
\mathrm{Var}_{\boldsymbol{T}} g(\boldsymbol{T}) &\approx \mathrm{E}_{\boldsymbol{T}} \left((g(\boldsymbol{T}) - g(\boldsymbol{\theta}))^2 \right) \\
&\approx \mathrm{E}_{\boldsymbol{T}} \left(\left(\sum_{i=1}^{k} g_i'(\boldsymbol{\theta})(t_i - \theta_i) \right)^2 \right) \\
&= \sum_{i=1}^{k} \{g_i'(\boldsymbol{\theta})\}^2 \mathrm{Var}_{\boldsymbol{T}} T_i + 2 \sum_{i>j} g_i'(\boldsymbol{\theta}) g_j'(\boldsymbol{\theta}) \mathrm{Cov}_{\boldsymbol{T}}(T_i, T_j)
\end{aligned}$$

Notes

Approx. Mean and Variance Example

Suppose X is a random variable with $\mathrm{E}(X) = \mu \neq 0$ and we want to estimate a function $g(\mu)$. A first order approximation gives

$$g(X) \approx g(\mu) + g'(\mu)(X - \mu).$$

If we use $g(X)$ as an estimator of $g(\mu)$ then

$$\begin{aligned}
\mathrm{E} g(X) &\approx g(\mu) \\
\mathrm{Var} g(X) &\approx \{g'(\mu)\}^2 \mathrm{Var}(X).
\end{aligned}$$

For example, if $g(\mu) = 1/\mu$ and we estimate $1/\mu$ with $1/X$, then

$$\begin{aligned}
\mathrm{E}\frac{1}{X} &\approx \frac{1}{\mu} \\
\mathrm{Var}\frac{1}{X} &\approx \left(\frac{1}{\mu}\right)^4 \mathrm{Var}(X)
\end{aligned}$$

Notes

Delta Method

Theorem:
Let Y_n be a sequence of random variables such that $\sqrt{n}(Y_n - \theta) \xrightarrow{D} \mathrm{N}(0, \sigma^2)$. For a given function g and a specific value of θ, suppose $g'(\theta) \neq 0$ exists. Then

$$\sqrt{n}\left\{g(Y_n) - g(\theta)\right\} \xrightarrow{D} \mathrm{N}\left(0, \sigma^2 \left\{g'(\theta)\right\}^2\right).$$

The proof uses Slutsky's Theorem. See CB for details.

Notes

Multivariate Delta Method

Theorem:
Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample with $\mathrm{E}(X_{ij}) = \mu_i$ and $\mathrm{Cov}(X_{ik}, X_{jk}) = \sigma_{ij}$. For a given function g with continuous first partial derivatives and a specific value of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)$ for which $\tau^2 = \sum\sum \sigma_{ij} \frac{\partial g(\boldsymbol{\mu})}{\partial \mu_i} \frac{\partial g(\boldsymbol{\mu})}{\partial \mu_j} > 0$,

$$\sqrt{n}\left\{g(\bar{X}_1, \ldots, \bar{X}_p) - g(\mu_1, \ldots, \mu_p)\right\} \xrightarrow{D} \mathrm{N}(0, \tau^2).$$

Notes

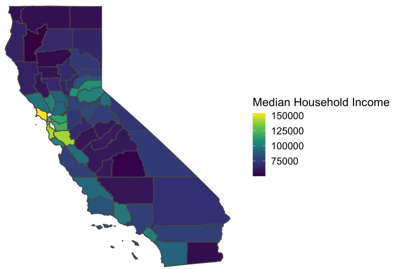
Case Study

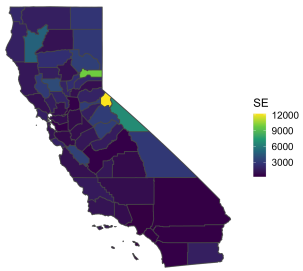
Data from the American Community Survey is often used to construct important population estimates for various domains. However, these estimates can often have high variance in domains with limited sample sizes.

Let's consider county level estimates of median household income for the state of California. If we denote the true median income for county i as μ_i, and the estimate as $\hat{\mu}_i$, then we might assume $\hat{\mu}_i \sim \mathrm{N}(\mu, \hat{\sigma}^2)$, where $\hat{\sigma}$ is the estimated standard error.

Notes

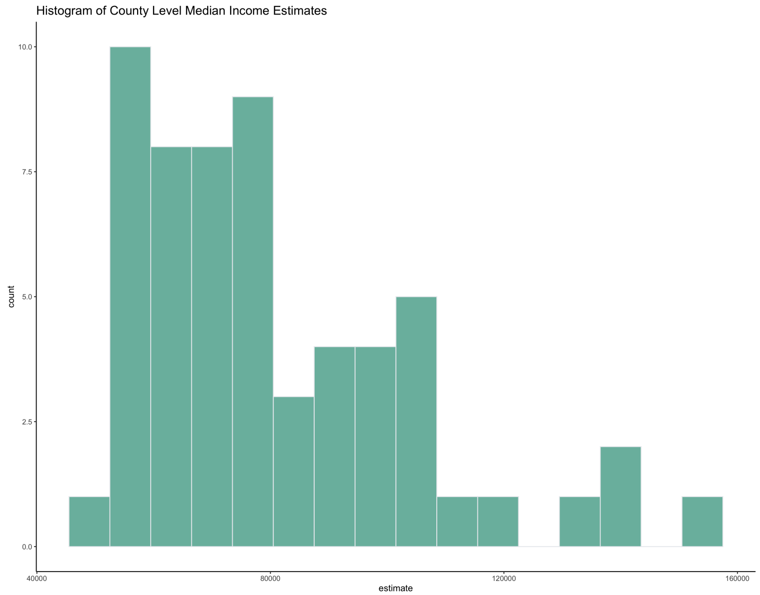
Case Study Cont.





Notes

Case Study Cont.



Notes

Case Study Cont.

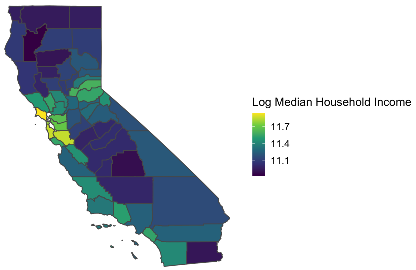
In order to improve these estimates, we might want to consider a statistical model (e.g. consider covariates, spatial correlation, etc.). Small area models use the data $(\hat{\mu}_i, \hat{\sigma})$, for $i = 1, \ldots, n$, to construct improved estimates with lower standard error. Because these estimates are skewed, it might be of interest to take a log transformation before modeling. However, when taking this transformation, we must consider the delta method to approximate the new standard error.

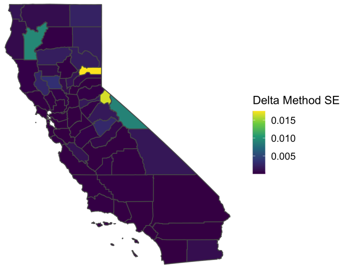
$$\mathrm{E}\log(\hat{\mu}_i) \approx \log(\mu_i)$$

$$\mathrm{Var}\log(\hat{\mu}_i) \approx \left(\frac{1}{\mu_i}\right)^2 \hat{\sigma}^2.$$

Notes

Case Study Cont.





Notes

Notes