Stat 135: Concepts of Statistics Lecture Notes

Joseph Pagadora
Instructor: Professor Yun Song

Fall 2017

Contents

Chapter 1

Sampling

Our first topic is sampling, in which we discuss the behavior and properties of estimates of population statistics through a sample (subset) of a population. Our general framework here is the following:
Suppose we our interested in some data which follows an unknown distribution P. We can obtain a sample $X_1, X_2, ..., X_n \sim P$, and given these datapoints, what can we learn from the unknown distribution?

1.1 Survey Sampling

Suppose $X_1, \dots, X_n$ are sampled from a distribution with unknown population-wide mean and variance. Also, we have:

(i) $X_1, \dots, X_n$ are **exchangeable** random variables. That is, if $F_{X_1,\dots,X_n}(x_1, \dots, x_n)$ is the c.d.f of the joint distribution of $X_1, \dots, X_n$, then for any permutation π of $x_1, \dots, x_n$, we have $F_{X_1,\dots,X_n}(x_1, \dots, x_n) = F_{X_1,\dots,X_n}(x_{\pi(1)}, \dots, x_{\pi(n)})$.

(ii) $X_1, \dots, X_n$ may or may not be independent.

Remark: Property (i) implies that $X_1, \dots, X_n$ are identically distributed.

Def: Given sample data $X_1, \dots, X_n$, the **sample mean** is $\overline{X}_n := \dfrac{X_1 + \cdots + X_n}{n}$.

Note that $\mathbb{E}(\overline{X}_n) = \mathbb{E}\Big[\frac{1}{n}(X_1 + \cdots + X_n)\Big] = \frac{1}{n}\mathbb{E}(X_1 + \cdots + X_n)$. If $X_1, \dots, X_n$ are exchangeable, and hence identically distributed, then for $\mathbb{E}(X_i) = \mu$, we have $\mathbb{E}(\overline{X}_n) = \frac{1}{n} \cdot n\mu = \mu$.

Def: An **estimator** $\hat{\theta}(X_1, \dots, X_n)$ for a parameter θ is said to be **unbiased** if $\mathbb{E}[\hat{\theta}(X_1, \dots, X_n)] = \theta$ for all values of "valid" θ and for all n.

1.1.1 Mean Square Error of an Estimator

This is defined to be $\mathbb{E}[(\hat{\theta}-\theta)^2]$. A very useful and important property of the mean square error is that it can be decomposed into variance and bias squared:

$$\mathbb{E}[(\hat{\theta}-\theta)^2] = \mathbb{E}\Big[\Big((\hat{\theta}-\mathbb{E}(\hat{\theta})) + (\mathbb{E}(\hat{\theta})-\theta)\Big)^2\Big] =$$

$$= \mathbb{E}\Big[(\hat{\theta}-\mathbb{E}(\hat{\theta}))^2\Big] + 2\mathbb{E}\Big[(\hat{\theta}-\mathbb{E}(\hat{\theta}))(\mathbb{E}(\hat{\theta})-\theta)\Big] + \mathbb{E}\Big[(\mathbb{E}(\hat{\theta})-\theta)^2\Big] =$$

$$= \mathbb{E}\Big[(\hat{\theta}-\mathbb{E}(\hat{\theta}))^2\Big] + 2(\mathbb{E}(\hat{\theta})-\theta)\underbrace{\cancelto{0}{\mathbb{E}(\hat{\theta}-\mathbb{E}(\hat{\theta}))}}_{0} + (\mathbb{E}(\hat{\theta})-\theta)^2 = \underbrace{\mathbb{E}\Big[(\hat{\theta}-\mathbb{E}(\hat{\theta}))^2\Big]}_{\text{variance of } \hat{\theta}} + \underbrace{(\mathbb{E}(\hat{\theta})-\theta)^2}_{\text{bias}^2}.$$

1.1.2 Sample Variance

Def: The **sample variance** of a sample of size n is $S_n^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$.

Observe that we divide by $n-1$ instead of n. This is because we want the sample variance to be an unbiased estimator for the variance σ^2. To prove this, it is helpful to first prove the following useful proposition.

Proposition: $\mathbb{E}\Big[\displaystyle\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\Big] = n\sigma^2 - n\text{Var}(\overline{X}_n)$.

Proof: We have

$$\mathbb{E}\Big[\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2)\Big] = \mathbb{E}\Big[\Big(\sum_{i=1}^{n}X_i^2\Big) - 2n\overline{X}_n\cdot\overline{X}_n + n\overline{X}_n^2\Big] =$$

$$= \sum_{i=1}^{n}\mathbb{E}(X_i^2) - n\mathbb{E}(\overline{X}_n^2) = \sum_{i=1}^{n}[\mu^2+\sigma^2] - n[\text{Var}(\overline{X}_n) + (\mathbb{E}(\overline{X}_n))^2] =$$

$$= n(\mu^2+\sigma^2) - n\text{Var}(\overline{X}_n) - n\mu^2 = n\sigma^2 - n\text{Var}(\overline{X}_n).$$

$\blacksquare$

Recall that we had two cases: either our sample $X_1,\dots,X_n$ were independent or not. First, let us consider the case in which they are independent. Then

$$\text{Var}(\overline{X}_n) = \frac{1}{n^2}\text{Var}(X_1+\cdots+X_n) = \frac{1}{n^2}\cdot n\text{Var}(X_1) = \frac{1}{n}\sigma^2$$

$$\implies \mathbb{E}\Big[\sum_{i=1}^{n}(X_i-\overline{X}_n)^2\Big] = n\sigma^2 - n\text{Var}(\overline{X}_n) = n\sigma^2 - \sigma^2 = \sigma^2(n-1).$$

Since we want $\mathbb{E}(S^2) = \sigma^2$, it easily follows from above why we divide by $n-1$.

1.2 Estimators for Normally-Distributed Data

Theorem: Suppose $X_1, \ldots, X_n$ are i.i.d. Normal(μ, σ^2). Then:

(i) $\overline{X}_n$ and $\displaystyle\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ are independent.

(ii) $\overline{X}_n \sim$Normal$\Big(\mu, \frac{\sigma^2}{n}\Big)$.

(iii) $\dfrac{\sum_{i=1}^{n}(X_i - \overline{X}_n)^2}{\sigma^2} \sim \chi^2_{n-1}$ (chi-square distribution with $n-1$ degrees of freedom).

This theorem gives us incredibly important information on the distribution of the sample mean and the sample variance of normally distributed data. However, in order to prove this theorem, we will need some lemmas. We will have to use the following fact to start, whose proof is left as an exercise (use change of variables technique):

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be continuous random variables with joint probability density function $f_{\mathbf{X}}$. Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ be defined by $\mathbf{Y} = A\mathbf{X}$, where A is a nonsingular real matrix. Then, the joint density of $\mathbf{Y}$ is given by $f_{\mathbf{Y}}(\mathbf{y}) = \dfrac{1}{|\det(A)|} \cdot f_{\mathbf{X}}(A^{-1}\mathbf{y})$.

Def: A symmetric real matrix $Q \in \mathbb{R}^{n\times n}$ is **orthogonal** if $QQ^T = Q^TQ = I$, the $n \times n$ identity matrix. (Note that $Q^{-1} = Q^T$.)

Lemma: If $Q \in \mathbb{R}^{n\times n}$ is orthogonal, then $|\det(Q)| = 1$.
Proof: Recall that for any matrix A, $\det(A) = \det(A^T)$. Then it is easy to see that $\det(Q^TQ) = \det(Q)^2$. But since $Q^TQ = I$, it easily follows that $|\det(Q)| = 1$. ■

Lemma: Suppose $Z_1, \ldots, Z_n$ are i.i.d. Normal$(0, 1)$ random variables, and suppose $\mathbf{Y} = Q\mathbf{Z}$, where Q is orthogonal. Then $Y_1, \ldots, Y_n$ are also i.i.d. Normal$(0, 1)$.
Proof: For each $1 \le i \le n$, $f_{Z_i}(z) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$.

So $f_{Z_1,\ldots,Z_n}(z_1, \ldots, z_n) = \dfrac{1}{(2\pi)^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}(z_1^2+\cdots+z_n^2)} = \dfrac{1}{(2\pi)^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}\mathbf{z}^T\mathbf{z}}$.

Since Q is orthogonal and hence invertible, it follows from the first lemma that we have

$$f_Y(\mathbf{y}) = \frac{1}{|\det(Q)|} f_Z(Q^{-1}\mathbf{y}) = f_Z(Q^T\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}(Q^T\mathbf{y})^TQ^T\mathbf{y}} = \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}\mathbf{y}^TQQ^T\mathbf{y}} =$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}\mathbf{y^Ty}} = \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot e^{-\frac{1}{2}(y_1^2+\cdots+y_n^2)}.$$

Thus $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim}$Normal$(0, 1)$. ■

<u>**Proof of the Theorem**</u>:

(i) It is easy to see that by the Gram-Schmidt Theorem, we can construct an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ whose first row is precisely $\left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right)$. Let $Z_1, \dots, Z_n$ be i.i.d Normal$(0,1)$ random variables. Suppose $\mathbf{Y} = Q\mathbf{Z}$.
Then $\mathbf{Y}^T\mathbf{Y} = \mathbf{Z}^TQ^TQ\mathbf{Z} = \mathbf{Z}^T\mathbf{Z}$.
Note also that $\mathbf{Y}^T\mathbf{Y} = \sum_{i=1}^n Y_i^2$, and that $Y_1 = \frac{1}{\sqrt{n}}(Z_1 + \cdots + Z_n) = \frac{n}{\sqrt{n}}\overline{Z}_n$.
This implies that $\overline{Z}_n = Y_1/\sqrt{n}$. Observe that

$$\sum_{i=1}^n (Z_i - \overline{Z}_n)^2 = \left(\sum_{i=1}^n Z_i^2\right) - n\overline{Z}_n^2 = \mathbf{Z}^T\mathbf{Z} - n\overline{Z}_n^2 = \mathbf{Y}^T\mathbf{Y} - Y_1^2 = \sum_{i=2}^n Y_i^2$$

By the second lemma, we know that $Y_1, \dots, Y_n$ are all independent, hence $\overline{Z}_n$ and $\sum_{i=1}^n (Z_i - \overline{Z}_n)^2$ are independent.
Now let $X_1, \dots, X_n$ be i.i.d. Normal(μ, σ^2). Define $Z_i = \dfrac{X_i - \mu}{\sigma}$.
Then $\overline{Z}_n = \dfrac{\overline{X}_n - \mu}{\sigma}$, and $\displaystyle\sum_{i=1}^n (Z_i - \overline{Z}_n)^2 = \sum_{i=1}^n \frac{(X_i - \overline{X}_n)^2}{\sigma^2}$. From this, we have proved (i).

(ii) This follows from the fact that linear combinations of normal variables are also normal. It thus remains to determine the parameters (mean and variance). It is easy to calculate that $\mathbb{E}[\overline{X}_n] = \mu$, and $\mathrm{Var}[\overline{X}_n] = \sigma^2/n$.

(iii) From part (i) above, we have $\displaystyle\sum_{i=1}^n \frac{(X_i - \overline{X}_n)^2}{\sigma^2} = Y_2^2 + \cdots + Y_n^2$. Recall that the sum of m i.i.d. squared standard normal random variables will have the χ_m^2 distribution. Thus, we have $\displaystyle\sum_{i=1}^n \frac{(X_i - \overline{X}_n)^2}{\sigma^2} \sim \chi_{n-1}^2$.

1.3 Sampling from a Finite Population

Suppose $X_1, \dots, X_n$ are sampled without replacement from a finite population of size N. Suppose this population has K distinct values $c_1, c_2, \dots, c_K$, where N_j members of the population take the value c_j, for $1 \le j \le K$.
Then $\mathbb{P}(X_i = c_a) = \dfrac{N_a}{N}$ for any $1 \le a \le K$ and for all i.
Then $\displaystyle\mathbb{E}(X_i) = \sum_{a=1}^K \frac{c_aN_a}{N} = \mu$, and $\displaystyle\mathrm{Var}(X_i) = \sum_{a=1}^K c_a^2 \cdot \frac{N_a}{N} - \mu^2 = \sigma^2$.
Furthermore, it is easy to see that $\mathbb{E}(\overline{X}_n) = \mu$. The variance calculation is non-trivial:

$$\mathrm{Var}(\overline{X}_n) = \mathrm{Var}\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big) = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \mathrm{Cov}(X_i, X_j).$$

For $i = j$, we have $\operatorname{Cov}(X_i, X_j) = \operatorname{Cov}(X_i, X_i) = \operatorname{Var}(X_i)$. For $i \neq j$, we have $\operatorname{Cov}(X_i, X_j) = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) = \mathbb{E}(X_i X_j) - \mu^2$. We have

$$\mathbb{E}(X_i X_j) = \sum_{a=1}^{K}\sum_{b=1}^{K} c_a c_b \mathbb{P}(X_i = a,\ X_j = c_b).$$

By exchangeability of $X_1, \ldots, X_n$, $\mathbb{P}(X_i = c_a,\ X_j = c_b) = \mathbb{P}(X_1 = c_a,\ X_2 = c_b)$.
If $a = b$, then

$$\mathbb{P}(X_1 = c_a,\ X_2 = c_b) = \frac{N_a}{N} \cdot \frac{N_a - 1}{N - 1},$$

and if $a \neq b$,

$$\mathbb{P}(X_1 = c_a,\ X_2 = c_b) = \frac{N_a}{N} \cdot \frac{N_b}{N - 1}.$$

Then

$$\mathbb{E}(X_i X_j) = \frac{1}{N(N-1)} \sum_{a=1}^{K} c_a^2 N_a (N_a - 1) + \frac{1}{N(N-1)} \sum_{a \neq b}^{K} c_a c_b N_a N_b.$$

Therefore

$$\operatorname{Var}(\overline{X}_n) = \frac{1}{n^2}\Big(\sum_{i \neq j}^{n}\Big[\sum_{a=1}^{K} \frac{c_a^2 N_a^2}{N(N-1)} - \frac{(\sigma^2 + \mu^2)}{N-1} + \frac{1}{N(N-1)} \sum_{a \neq b}^{K} c_a c_b N_a N_b - \mu^2\Big] + n\sigma^2\Big) =$$

$$= \frac{n-1}{n}\Big[\frac{1}{N(N-1)}\Big(\sum_{a=1}^{K} c_a^2 N_a^2 + \sum_{a \neq b}^{K} c_a c_b N_a N_b\Big) - \mu^2 - \frac{(\sigma^2 + \mu^2)}{N-1}\Big] + \frac{\sigma^2}{n} =$$

$$= \frac{\sigma^2}{n} - \frac{n-1}{n}\mu^2 - \frac{n-1}{n} \cdot \frac{\sigma^2 + \mu^2}{N-1} + \frac{n-1}{n} \cdot \frac{1}{N(N-1)} \sum_{a=1}^{K}\sum_{b=1}^{K} c_a c_b N_a N_b =$$

$$= \frac{\sigma^2}{n}\Big(1 - \frac{n-1}{N-1}\Big) - \frac{n-1}{n}\mu^2\Big(1 + \frac{1}{N-1}\Big) + \frac{n-1}{n} \cdot \frac{1}{N(N-1)}\Big(\sum_{a=1}^{K} c_a N_a\Big)^2 =$$

$$= \frac{\sigma^2}{n}\Big(\frac{N-n}{N-1}\Big) - \frac{n-1}{n} \cdot \frac{N}{N-1}\mu^2 + \frac{n-1}{n} \cdot \frac{N}{N-1}\mu^2 = \frac{\sigma^2}{n}\Big(\frac{N-n}{N-1}\Big).$$

This gives the following result:

$$\mathbb{E}\Big[\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\Big] = n\sigma^2 - n \operatorname{Var}(\overline{X}_n) = n\sigma^2 - n \cdot \frac{\sigma^2}{n}\Big(\frac{N-n}{N-1}\Big) = (n-1)\Big(\frac{N}{N-1}\Big)\sigma^2$$

for $X_1, \ldots, X_n$ sampled without replacement from a population of size N.
Recall the sample variance is $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$, so we see that

$$\mathbb{E}\Big[\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\Big] = (n-1)S^2 = (n-1)\Big(\frac{N}{N-1}\Big)\sigma^2, \text{ so } \sigma^2 = \frac{N-1}{N}S^2.$$

Therefore, an unbiased estimator of σ^2 is $\Big(1 - \frac{1}{N}\Big)S^2$, and an unbiased estimator of $\operatorname{Var}(\overline{X}_n)$ is $\Big(1 - \frac{n}{N}\Big)\frac{S^2}{n}$.

1.4 Approximating the Samping Distribution of $\overline{X}_n$

Here we recall the Central Limit Theorem to approximate the distribution of our sample mean $\overline{X}_n$, assuming a sufficiently large population, regardless of how the data is distributed. This approximation is especially useful to construct approximate confidence intervals.

Def: Let X be a random variable with c.d.f. F. Let (X_n) be a sequence of random variables, each with c.d.f. F_n. We say that X_n **converges in distribution** to X, denoted $X_n \xrightarrow{d} X$, if $F_n(x)$ converges pointwise to $F(x)$ for all x at which F is continuous.

Central Limit Theorem:
If $X_1, \ldots, X_n$ are i.i.d. with $\mathbb{E}(X_i) = \mu$, and $\mathrm{Var}(X_i) = \sigma^2 < \infty$, then

$$\sqrt{n}\Big(\frac{\overline{X}_n - \mu}{\sigma}\Big) \to Z \sim \mathrm{Normal}(0,1) \text{ as } n \to \infty.$$

Def: For $0 < \alpha < 1$, a 100$(1-\alpha)$% **confidence interval** for a parameter θ is a random interval $I(X_1, \ldots, X_n)$ such that $\mathbb{P}(\theta \in I(X_1, \ldots, X_n)) = 1 - \alpha$.

Ex: Let $z(\alpha)$ denote the point at which $\mathbb{P}(Z > z(\alpha)) = \alpha$, where Z is a standard normal random variable. By symmetry of Z about 0, we have $\mathbb{P}\Big(-z(\alpha/2) \le Z \le z(\alpha)/2\Big) = 1 - \alpha$. So, by the Central Limit Theorem,

$$\mathbb{P}\Big(-z(\alpha/2) \le \frac{\overline{X}_n - \mu}{\sqrt{\mathrm{Var}(\overline{X}_n)}} \le z(\alpha/2)\Big) \approx 1 - \alpha.$$

This means that with probability $1 - \alpha$, μ lies in the interval

$$\Big(\overline{X}_n - z(\alpha/2)\sqrt{\mathrm{Var}(\overline{X}_n)},\ \ \overline{X}_n + z(\alpha/2)\sqrt{\mathrm{Var}(\overline{X}_n)}\Big).$$

1.5 Summary of Results

Sampling with Replacement	**Sampling without Replacement**
$\mathbb{E}(\overline{X}_n) = \mu$	$\mathbb{E}(\overline{X}_n) = \mu$
$\mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n}$	$\mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n}\Big(\frac{N-n}{N-1}\Big)$
$\mathbb{E}\Big[\sum_{i=1}^n (X_i - \overline{X}_n)^2\Big] = n\sigma^2 - n\ \mathrm{Var}(\overline{X}_n)$	$\mathbb{E}\Big[\sum_{i=1}^n (X_i - \overline{X}_n)^2\Big] = n\sigma^2 - n\ \mathrm{Var}(\overline{X}_n)$
$\mathbb{E}(S_n^2) = \sigma^2$	$\mathbb{E}(S_n^2) = \sigma^2\Big(\frac{N}{N-1}\Big)$
S_n^2 is an unbiased estimator of σ^2	$\Big(1 - \frac{1}{N}\Big)S_n^2$ is an unbiased estimator of σ^2
$\frac{S_n^2}{n}$ is an unbiased estimator for $\mathrm{Var}(\overline{X}_n)$	$\frac{S_n^2}{n}\Big(1 - \frac{n}{N}\Big)$ is an unbiased estimator for $\mathrm{Var}(\overline{X}_n)$

Chapter 2

Fitting Parametric Distributions