Parameter Estimation

Eklavya Sharma

Our aim is to find out something about a distribution by observing a sample.

Definition 1 (Sample). *For a distribution D, a* sample of size n *from D is the sequence $[X_1, X_2, \ldots, X_n]$ of n IID random variables, each having distribution D.*

Notation: For a random variable X having distribution D and any function g, define $\mathrm{E}(g(D)) := \mathrm{E}(g(X))$. (Hence, $\mathrm{Var}(D) := \mathrm{Var}(X)$.)

1 Bias and Variance of Estimators

Definition 2 (Sample mean and variance). *Let $[X_1, \ldots, X_n]$ be a sample.*

1. *The mean of the sample is defined as $\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$.*
2. *The variance of the sample is defined as $V_X := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.*
3. *The standard-deviation of the sample is defined as $S_X := \sqrt{V_X}$.*

Theorem 1. *Let $\overline{X}$ be the mean of a sample from D. Then $\mathrm{E}(\overline{X}) = \mathrm{E}(D)$ and $\mathrm{Var}(\overline{X}) = \mathrm{Var}(D)/n$.*

Claim 2. *Let $\overline{X}$ and S^2 be the mean and variance, respectively, of sample $[X_1, \ldots, X_n]$. Let a be any random variable (or a constant). Then*

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}(X_i - a)^2 - n(\overline{X} - a)^2\right).$$

(Note that setting $a = \overline{X}$ gives the definition of S^2.)

Theorem 3. *Let V be the variance of sample $[X_1, \ldots, X_n]$ from D. Let $\mu := \mathrm{E}(D)$ and $\sigma^2 := \mathrm{Var}(D)$. Then $\mathrm{E}(V) = \sigma^2$ and $\mathrm{Var}(V) = \frac{\mathrm{E}((D-\mu)^4)}{n} - \frac{\sigma^4(n-3)}{n(n-1)}$.*

Proof.

$$\begin{aligned}
\mathrm{E}(V) &= \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathrm{E}((X_i - \mu)^2) - n\,\mathrm{E}((\overline{X} - \mu)^2)\right) && \text{(by Claim 2)} \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathrm{Var}(X_i) - n\,\mathrm{Var}(\overline{X})\right) = \sigma^2.
\end{aligned}$$

The expression for $\mathrm{Var}(V)$ is from [7]. □

2 Distribution of Estimators

Definition 3. *Let Z be a random variable and $S := [X_1, X_2, \ldots]$ be an infinite sequence of random variables. We say that S converges to Z if $\lim_{n\to\infty} F_{X_n}(x) = F_Z(x)$ for all $x \in \mathbb{R}$ where F_Z is continuous.*

Theorem 4 (Central Limit Theorem)**.** *Let $X_1, X_2, \ldots$ be IID randvars having mean μ and variance σ^2. Let $\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$. Let $Y_n := \sqrt{n/\sigma}(\overline{X}_n - \mu)$. Then $[Y_1, Y_2, \ldots]$ converges to $N(0, 1)$.*

Lemma 5 (Scaling normal)**.** *Let $X \sim N(\mu, \sigma)$. Then for any constants a and b, $aX + b \sim N(a\mu + b, |b|\sigma)$.*

Lemma 6 ([4])**.** *Let X and Y be independent randvars where $X \sim N(\mu_X, \sigma_X)$ and $Y \sim N(\mu_Y, \sigma_Y)$. Then $X + Y \sim N(\mu_X + \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2})$.*

Theorem 7. *Let $[X_1, \ldots, X_n]$ be a sample from $N(\mu, \sigma)$. Let $\overline{X}$ and S^2 be the mean and variance of the sample. Then*

1. $\overline{X} \sim N(\mu, \sigma/\sqrt{n})$.
2. $\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n - 1)$.
3. *$\overline{X}$ and S^2 are independent.*

Here $\chi^2(n - 1)$ is the Chi-Squared distribution with $n - 1$ degrees of freedom.

Proof. Part 1 follows from Lemmas 5 and 6.

[3] proves parts 2 and 3. Alternatively, [5] proves part 3 and [2] proves part 2. □

3 Distribution of Statistical Scores

Definition 4. *Let $Z \sim N(0, 1)$ and $U \sim \chi^2(r)$ be independent randvars. Let $T := Z/\sqrt{U/r}$. Then T's distribution is called the* Student's t distribution *with r degrees of freedom.*

Lemma 8 (t distribution is symmetric)**.** *Let $T \sim t(r)$. Then T and $-T$ have the same distribution.*

Proof. Let $Z \sim N(0, 1)$ and $U \sim \chi^2(r)$ be independent randvars and $T := Z/\sqrt{U/r}$. Then $T \sim t(r)$. Since $-Z \sim N(0, 1)$, so $-T = (-Z)/\sqrt{U/r} \sim t(r)$. □

Lemma 9 (Implications of symmetry)**.** *Let X be a continuous random variable such that X and $-X$ have the same distribution. Then, $\forall x \in \mathbb{R}$, we get $F_X(x) + F_X(-x) = 1$, and $\forall \alpha \in [0, 1]$, we get $F_X^{-1}(\alpha) + F_X^{-1}(1 - \alpha) = 0$.*

Proof. $F_X(-x) = F_{-X}(-x) = \Pr(-X \le -x) = \Pr(X \ge x) = 1 - F_X(x)$.

Let $x = F_X^{-1}(\alpha)$. Then
$-F_X^{-1}(1 - \alpha) = -F_X^{-1}(1 - F_X(x)) = -F_X^{-1}(F_X(-x)) = x = F_X^{-1}(\alpha)$. □

Theorem 10. *Let $\overline{X}$ and S^2 be the mean and variance of a sample from $N(\mu, \sigma)$. Then*

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

Proof sketch. Use Theorem 7 and $\dfrac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2/\sigma^2}{n-1}}} = \dfrac{\overline{X} - \mu}{S/\sqrt{n}}$. □

4 Distribution of Paired Statistical Scores

Theorem 11. *Let $\overline{X}$ and S_X^2 be the mean and variance of a sample $[X_1, \ldots, X_n]$ from distribution $N(\mu_X, \sigma)$. Let $\overline{Y}$ and S_Y^2 be the mean and variance of sample $[Y_1, \ldots, Y_m]$ from distribution $N(\mu_Y, \sigma)$. The two samples are independent. Then for*

$$S_p^2 := \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}, \qquad T := \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}},$$

we have $T \sim t(n+m-2)$. (S_p^2 is called pooled sample variance.*)*

Proof sketch. $\overline{X}, \overline{Y}, S_X, S_Y$ are independent by Theorem 7.3.

$$\overline{X} \sim N(\mu_X, \sigma/\sqrt{n}) \quad \text{and} \quad \overline{Y} \sim N(\mu_Y, \sigma/\sqrt{m}) \qquad \text{(by Theorem 7.1)}$$

$$\implies \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0,1). \qquad \text{(by Lemmas 5 and 6)}$$

$$(n-1)S_X^2/\sigma^2 \sim \chi^2(n-1) \quad \text{and} \quad (m-1)S_Y^2/\sigma^2 \sim \chi^2(m-1) \qquad \text{(by Theorem 7.2)}$$

$$\implies (n+m-2)S_p^2/\sigma^2 \sim \chi^2(n+m-2).$$ □

Lemma 12. *For $i \in \{1, \ldots, k\}$, let $\mathbf{X}_i := [X_{i,1}, \ldots, X_{i,n_i}]$ be a sample from $N(\mu_i, \sigma_i)$. The samples are independent. Let $a_1, \ldots, a_k$ be non-negative constants. Let S_i^2 be the variance of $\mathbf{X}_i$. Let*

$$r := \frac{\left(\sum_{i=1}^k a_i S_i^2\right)^2}{\sum_{i=1}^k \frac{(a_i S_i^2)^2}{n_i - 1}} \qquad L := \frac{r}{\sum_{i=1}^k a_i \sigma_i^2} \sum_{i=1}^k a_i S_i^2.$$

Then L is approximately distributed $\chi^2(r)$.

Proof. The meaning of *approximate* and the 'proof' can be found at [6, 8]. □

Theorem 13. *Let $\overline{X}$ and S_X^2 be the mean and variance of a sample $[X_1, \ldots, X_n]$ from distribution $N(\mu_X, \sigma_X)$. Let $\overline{Y}$ and S_Y^2 be the mean and variance of sample $[Y_1, \ldots, Y_m]$ from distribution $N(\mu_Y, \sigma_Y)$. The samples $[X_1, \ldots, X_n]$ and $[Y_1, \ldots, Y_m]$ are independent. Then for*

$$r := \frac{(S_X^2/n + S_Y^2/m)^2}{\frac{(S_X^2/n)^2}{n-1} + \frac{(S_Y^2/m)^2}{m-1}} \quad \text{and} \quad T := \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{S_X^2/n + S_Y^2/m}},$$

T approximately follows $t(r)$.

Proof sketch. $T = Z/(\sqrt{L/r})$, where

$$Z := \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\sigma_X^2/n + \sigma_Y^2/m}} \sim N(0,1), \qquad L := \frac{r}{\sigma_X^2/n + \sigma_Y^2/m}\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right),$$

and L approximately follows $\chi^2(r)$ by Lemma 12. □

Definition 5. *Let X and Y be independent randvars, where $X \sim \chi^2(u)$ and $Y \sim \chi^2(v)$. Then the distribution of $\frac{X/u}{Y/v}$ is called the F distribution with parameters u and v.*

Lemma 14. *Let R be an F distribution with parameters u and v. Then R^{-1} is an F distribution with parameters v and u. Furthermore, $\forall x \in \mathbb{R}_{>0}$, we get $F_R(x)+F_{R^{-1}}(x^{-1}) = 1$, and $\forall \alpha \in [0,1]$, we get $F_R^{-1}(\alpha)F_{R^{-1}}^{-1}(1-\alpha) = 1$.*

Proof. $F_{R^{-1}}(x^{-1}) = \Pr(R^{-1} \le x^{-1}) = \Pr(R \ge x) = 1 - F_R(x)$.

Let $x := F_R^{-1}(\alpha)$. Then
$F_{R^{-1}}^{-1}(1-\alpha) = F_{R^{-1}}^{-1}(1 - F_R(x)) = F_{R^{-1}}^{-1}(F_{R^{-1}}(x^{-1})) = x^{-1} = 1/F_R^{-1}(\alpha)$. □

5 Correlated Data

Let X and Y be random variables over a joint distribution D. Let $[(X_i, Y_i) : i \in \{1, \dots, n\}]$ be an IID sample drawn from D. Let $\overline{X} := (1/n)\sum_{i=1}^n X_i$ and $\overline{Y} := (1/n)\sum_{i=1}^n Y_i$.

Theorem 15. $\mathrm{Cov}(\overline{X}, \overline{Y}) = \mathrm{Cov}(X_1, Y_1)/n$.

Definition 6 (Sample covariance)**.**

$$C_{X,Y} := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y}).$$

Claim 16. *For any real numbers (or random variables) a and b,*

$$C_{X,Y} = \frac{1}{n-1}\left(\sum_{i=1}^n (X_i - a)(Y_i - b) - n(\overline{X} - a)(\overline{Y} - b)\right).$$

Theorem 17. $\mathrm{E}(C_{X,Y}) = \mathrm{Cov}(X_1, Y_1)$.

Proof. Let $\mu_X := \mathrm{E}(X_1)$ and $\mu_Y := \mathrm{E}(Y_1)$. Then

$$\begin{aligned}\mathrm{E}(C_{X,Y}) &= \frac{1}{n-1}\left(\sum_{i=1}^n \mathrm{E}((X_i - \mu_X)(Y_i - \mu_Y)) - n\,\mathrm{E}((\overline{X} - \mu_X)(\overline{Y} - \mu_Y))\right) \quad \text{(by Claim 16)}\\ &= \frac{1}{n-1}\left(\sum_{i=1}^n \mathrm{Cov}(X_i, Y_i) - n\,\mathrm{Cov}(\overline{X}, \overline{Y})\right) = \mathrm{Cov}(X_1, Y_1).\end{aligned}$$

□

Let V_X and V_Y be the sample variance of $[X_1, \dots, X_n]$ and $[Y_1, \dots, Y_n]$, respectively.

Lemma 18.

$$(n-1)^2(V_XV_Y - C_{X,Y}^2) = \sum_{1\le i<j\le n} ((X_i-\overline{X})(Y_j-\overline{Y}) - (X_j-\overline{X})(Y_i-\overline{Y}))^2.$$

Proof. Let $W_i := X_i - \overline{X}$ and $Z_i := Y_i - \overline{Y}$. Then

$$(n-1)^2(V_XV_Y - C_{X,Y}^2) = \left(\sum_{i=1}^n W_i^2\right)\left(\sum_{i=1}^n Z_i^2\right) - \left(\sum_{i=1}^n W_iZ_i\right)^2$$

$$= \sum_{i=1}^n\sum_{j=1}^n (W_i^2Z_j^2 - W_iZ_iW_jZ_j) = \sum_{i=1}^n\sum_{j=1}^n W_iZ_j(W_iZ_j - W_jZ_i)$$

$$= \sum_{1\le i<j\le n} (W_iZ_j - W_jZ_i)^2.$$

□

Lemma 19.

$$\mathrm{E}(V_XV_Y - C_{X,Y}^2) = \frac{n}{n-1}\left(\mathrm{Var}(X_1)\,\mathrm{Var}(Y_1) - \mathrm{Cov}(X_1,Y_1)\right).$$

Proof. Let $W_i := X_i - \overline{X}$ and $Z_i := Y_i - \overline{Y}$. Then

$$(n-1)^2\,\mathrm{E}(V_XV_Y - C_{X,Y}^2) = \sum_{1\le i<j\le n} \mathrm{E}((W_iZ_j - W_jZ_i)^2)$$

$$= \sum_{1\le i<j\le n} (\mathrm{E}(W_i^2Z_j^2) + \mathrm{E}(W_j^2Z_i^2) - 2\,\mathrm{E}(W_iW_jZ_iZ_j))$$

$$= \sum_{1\le i<j\le n} (\mathrm{E}(W_i^2)\,\mathrm{E}(Z_j^2) + \mathrm{E}(W_j^2)\,\mathrm{E}(Z_i^2) - 2\,\mathrm{E}(W_iZ_i)\,\mathrm{E}(W_jZ_j))$$

(samples are independent)

$$= \sum_{1\le i<j\le n} (\mathrm{Var}(X_1)\,\mathrm{Var}(Y_1) + \mathrm{Var}(X_1)\,\mathrm{Var}(Y_1) - 2\,\mathrm{Cov}(X_1,Y_1)^2)$$

$$= n(n-1)(\mathrm{Var}(X_1)\,\mathrm{Var}(Y_1) - \mathrm{Cov}(X_1,Y_1)^2).$$

□

6 Linear Regression

Let $[(x_i,y_i) : i \in \{1,\ldots,n\}]$ be our data. Define

$$\mathrm{SSE}(\alpha,\beta) := \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2.$$

- Let $\overline{X}$ and V_X be the mean and variance of sample $[x_1,\ldots,x_n]$.
- Let $\overline{Y}$ and V_Y be the mean and variance of sample $[y_1,\ldots,y_n]$.
- Let $C_{X,Y}$ be the covariance of sample $[(x_i,y_i) : i \in \{1,\ldots,n\}]$.

Theorem 20.

$$\mathrm{SSE}(\alpha,\beta) = n(\alpha + \beta\overline{X} - \overline{Y})^2 + (n-1)V_X\left(\beta - \frac{C_{X,Y}}{V_X}\right)^2 + (n-1)\left(V_Y - \frac{C_{X,Y}^2}{V_X}\right).$$

Hence, SSE is minimized at $(\widehat{\alpha},\widehat{\beta})$, where $\widehat{\beta} := C_{X,Y}/V_X$ and $\widehat{\alpha} := \overline{Y} - \widehat{\beta}\overline{X}$.

6.1 Independent errors with mean 0

Let $[x_1, \ldots, x_n]$ be constants and $[e_1, \ldots, e_n]$ be IID random variables. For all $i \in \{1, \ldots, n\}$, let $\mathrm{E}(e_i) = 0$, $\mathrm{Var}(e_i) = \sigma^2$, and $y_i := \alpha + \beta x_i + e_i$.

- Let $\overline{x}$ and V_X be the mean and variance of sample $[x_1, \ldots, x_n]$.
- Let $\overline{y}$ and V_Y be the mean and variance of sample $[y_1, \ldots, y_n]$.
- Let $\overline{e}$ and V_E be the mean and variance of sample $[e_1, \ldots, e_n]$.
- Let $C_{X,Y}$ be the covariance of sample $[(x_i, y_i) : i \in \{1, \ldots, n\}]$.
- Let $C_{X,E}$ be the covariance of sample $[(x_i, e_i) : i \in \{1, \ldots, n\}]$.

Lemma 21.

$$C_{X,E} = \sum_{i=1}^{n} \frac{x_i - \overline{x}}{n-1} e_i \qquad \overline{y} = \alpha + \beta\overline{x} + \overline{e}$$

$$C_{X,Y} = \beta V_X + C_{X,E} \qquad V_Y = \beta^2 V_X + V_E + 2\beta C_{X,E}$$

Lemma 22.

$$\widehat{\beta} = \beta + \frac{C_{X,E}}{V_X} \qquad \mathrm{SSE}(\widehat{\alpha}, \widehat{\beta}) = (n-1)\left(V_E - \frac{C_{X,E}^2}{V_X}\right)$$

For any $t \in \mathbb{R}$,

$$\widehat{\alpha} + \widehat{\beta}t = (\alpha + \beta t) + \overline{e} + \frac{t - \overline{x}}{V_X} C_{X,E} = (\alpha + \beta t) + \sum_{i=1}^{n} \left(\frac{1}{n} + \frac{(t - \overline{x})(x_i - \overline{x})}{V_X(n-1)}\right) e_i.$$

Lemma 23. $\mathrm{E}(C_{X,E}) = 0$ *and* $\mathrm{Var}(C_{X,E}) = \mathrm{E}(C_{X,E}^2) = \sigma^2 V_X/(n-1)$.

Lemma 24. $\mathrm{E}(\mathrm{SSE}(\widehat{\alpha}, \widehat{\beta})) = (n-2)\sigma^2$.

Lemma 25. *For any $t \in \mathbb{R}$, $\mathrm{E}(\widehat{\beta}) = \beta$, $\mathrm{E}(\widehat{\alpha} + \widehat{\beta}t) = \alpha + \beta t$,*

$$\mathrm{Var}(\widehat{\beta}) = \frac{\sigma^2}{(n-1)V_X}, \qquad \mathrm{Var}(\widehat{\alpha} + \widehat{\beta}t) = \sigma^2\left(\frac{1}{n} + \frac{(t - \overline{x})^2}{(n-1)V_X}\right).$$

Lemma 26. $\mathrm{Cov}(\overline{y}, \widehat{\beta}) = \mathrm{Cov}(\overline{e}, C_{X,E}) = 0$.

Proof.

$$\mathrm{Cov}(\overline{y}, \widehat{\beta}) = \frac{\mathrm{Cov}(\overline{e}, C_{X,E})}{V_X} = \frac{\mathrm{Cov}(n\overline{e}, (n-1)C_{X,E})}{n(n-1)V_X}.$$

$$\mathrm{Cov}(n\overline{e}, (n-1)C_{X,E}) = \mathrm{Cov}\left(\sum_{i=1}^{n} e_i, \sum_{j=1}^{n} (x_j - \overline{x}) e_j\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(e_i, (x_j - \overline{x})e_j) = \sum_{i=1}^{n} (x_i - \overline{x})\sigma^2 = 0. \qquad \square$$

6.2 IID normal errors with mean 0

Let $[e_1, \ldots, e_n]$ be drawn IID from $N(0, \sigma)$.

Lemma 27.

$$\overline{y} \sim N\left(\alpha + \beta\overline{x}, \frac{\sigma}{\sqrt{n}}\right) \qquad \widehat{\beta} \sim N\left(\beta, \frac{\sigma}{\sqrt{(n-1)V_X}}\right)$$

Lemma 28. *$\overline{y}$, $\widehat{\beta}$, and* $\mathrm{SSE}(\widehat{\alpha}, \widehat{\beta})$ *are independent.* $\mathrm{SSE}(\widehat{\alpha}, \widehat{\beta})/\sigma^2 \sim \chi^2(n-2)$.

Proof. [1] gives a proof idea and lists references to proofs. □

Lemma 29.

$$\sqrt{\frac{(n-2)(n-1)V_X}{\mathrm{SSE}(\widehat{\alpha}, \widehat{\beta})}}(\widehat{\beta} - \beta) \sim t_{n-2}.$$

References

[1] Confidence intervals for regression parameters. Penn State STAT 415. URL: https://online.stat.psu.edu/stat415/lesson/7/7.5.

[2] Sampling distribution of sample variance. Penn State STAT 414. URL: https://online.stat.psu.edu/stat414/lesson/26/26.3.

[3] Show $\overline{X}$ and S^2 are independent. Duke STA 611 Fall 2019 Lecture 12. URL: http://www2.stat.duke.edu/courses/Fall19/sta611.01/Lecture/lec12_mean_var_indep.pdf.

[4] Sum of normally distributed random variables. Wikipedia. URL: https://en.wikipedia.org/wiki/Sum_of_normally_distributed_random_variables.

[5] guy. Proof of the independence of the sample mean and sample variance. Mathematics Stack Exchange. URL: https://math.stackexchange.com/q/50598.

[6] DGK. Proof and precise formulation of Welch-Satterthwaite equation. Mathematics Stack Exchange. URL: https://math.stackexchange.com/q/3189589.

[7] user940. Variance of sample variance? Mathematics Stack Exchange, October 2011. URL: https://math.stackexchange.com/q/73080.

[8] B. L. Welch. The generalization of 'student's' problem when several different population variances are involved. *Biometrika*, 34(1/2):28–35, 1947. doi:10.2307/2332510.