STAT24510 (W25): Statistical Theory and Methods IIa

Lecturer: Mei Wang

Notes by: Aden Chen

Thursday 16th January, 2025

Contents

1 Introduction

The goal of statistics is often to estimate a (population) parameter θ. From data, we may obtain point estimates $\hat{\theta}$ that depends on data, and construct confidence intervals to quantify the uncertainty of the estimate, with which we can conduct hypothesis testing.

In this course we will start from confidence intervals and hypothesis testing, and then move on to linear models. This course will be less structured; intend, it will be more like a collection of methods/producers.

1.1 Examples of Construction: Review of Wald and Wilson

1.1.1 Pivotal method

Example 1.1 (Pivotal method, Normal, Known σ^2). Let $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, 5^2)$, $i = 1, \ldots, n$. Our goal is to construct a $1 - \alpha$ CI for μ. We have the MLE estimator $\hat{\mu} = \overline{X} = n^{-1} \sum_{i=1}^{n} X_i$.

Note that

$$\frac{\overline{X} - \mathrm{E}[X]}{\sqrt{\mathrm{Var}(X)}} = \frac{\overline{X} - \mu}{\sqrt{5^2/n}} \sim \mathcal{N}(0, 1).$$

In particular, note that the left side is a function of data and parameters, while the right side is free of parameters.

We thus have

$$\mathbb{P}\left(z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\sqrt{5^2/n}} \leq z_{1-\alpha/2} \right) = 0.95,$$

using which we can construct the CI of μ: With $I := \left[\overline{X} - z_{1-\alpha/2}\sqrt{5^2/n}, \overline{X} + z_{1-\alpha/2}\sqrt{5^2/n} \right]$, we have $\mathbb{P}\left(\mu \in I \right) = 1 - \alpha$.

Notice that we obtain a probability statement of random interval containing a fixed quantity from a probability statement of a fixed interval containing a random quantity.

Example 1.2 (Pivotal method, Normal, Known μ). Let $X_i \overset{\text{iid}}{\sim} \mathcal{N}(4, \sigma^2)$, $i = 1, \ldots, n$. The goal: CI for σ^2, i.e., to find random variables L and U such that $\mathbb{P}\left(L \leq \sigma^2 \leq U \right) = 1 - \alpha$.

Note that

$$Y_i := \frac{X_i - 4}{\sigma} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$$

and thus

$$T_n := \sum Y_i^2 = \sum \left(\frac{X_i - 4}{\sigma} \right)^2 \sim \chi_n^2.$$

Again, we obtained a function of data and parameters that follows a known distribution. From

$$\mathbb{P}\left(\chi_{n,\alpha/2}^2 \leq T_n \leq \chi_{n,1-\alpha/2}^2 \right) = 1 - \alpha$$

we may again obtain the CI for σ^2,

$$\left[\frac{\sum (X_i - 4)^2}{\chi_{n,1-\alpha/2}^2}, \frac{\sum (X_i - 4)^2}{\chi_{n,\alpha/2}^2} \right].$$

3

Example 1.3 (Pivot failing). Let $X_i \overset{\text{iid}}{\sim}$ Bernoulli(p). The goal: CI for p. The MLE for p is $\hat{p} = \overline{X}$, thus we may be temped to try

$$T_n := \frac{\hat{p} - p}{\sqrt{\text{Var}(\hat{p})}},$$

but the distribution of T_n depends on p. The method of pivots fail.

1.1.2 Asymptotic CI

I.e., when we have large sample size n.

Example 1.4 (Wald CI). Let X_i be iid with mean μ and variance σ^2. From the CLT we have

$$\sqrt{n}\left(\overline{X} - \mu\right) \overset{\mathcal{D}}{\longrightarrow} \mathcal{N}(0, \sigma^2),$$

that is,

$$\lim_{n\to\infty} \mathbb{P}(T_n \leq x) = P(Z \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt.$$

Thus we have

$$\mathbb{P}\left(z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}} \leq z_{1-\alpha/2}\right) \approx 1 - \alpha.$$

When σ^2 is known, we may derive an approximate CI for μ:

$$\mathbb{P}\left(\overline{X} - z_{1-\alpha/2}\sqrt{\sigma^2/n} \leq \mu \leq \overline{X} + z_{1-\alpha/2}\sqrt{\sigma^2/n}\right) \approx 1 - \alpha.$$

When σ^2 is unknown: If there exists random variables $U_n \to_{\mathrm{p}} \sigma^2$ (that is, $\lim \mathbb{P}\left(U_n = \sigma^2\right) = 1$), then

$$T_n := \frac{\overline{X} - \mu}{\sqrt{U_n/n}} = \frac{\frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}}}{\sqrt{U_n/\sigma^2}}$$

where $(\overline{X} - \mu)/\sqrt{\sigma^2/n} \overset{\mathcal{D}}{\longrightarrow} \mathcal{N}(0,1)$ and $\sqrt{U_n/\sigma^2} \to_{\mathrm{p}} 1$, and thus by Slutsky's theorem we have

$$T_n \overset{\mathcal{D}}{\longrightarrow} \mathcal{N}(0,1),$$

using which we can again construct an approximate CI. Note that we used asymptotic approximation multiple times. This is called the Wald confidence interval.

Example 1.5 (Wald CI). Let $X_i \overset{\text{iid}}{\sim}$ Poisson(λ). The goal: asymptotic CI for λ. Note that we have the MLE of λ, $\hat{\lambda} = \overline{X}$, with $\mathrm{E}[\hat{\lambda}] = \lambda$ and $\text{Var}[\hat{\lambda}] = \lambda/n$. We then have

$$\frac{\overline{X} - \lambda}{\sqrt{\lambda/n}} \approx Z \sim \mathcal{N}(0,1).$$

We approximate a second time: $\frac{\overline{X} - \lambda}{\sqrt{\hat{\lambda}/n}} \approx Z$, from which we obtain the Wald CI for λ:

$$\left[\hat{\lambda} - z_{1-\alpha/2}\sqrt{\hat{\lambda}/n}, \hat{\lambda} + z_{1-\alpha/2}\sqrt{\hat{\lambda}/n}\right].$$

Example 1.6 (Wilson's method). Assume the same setup as above. Again, we use

$$\frac{\overline{X} - \lambda}{\sqrt{\lambda/n}} \approx Z \sim \mathcal{N}(0, 1),$$

which gives

$$\mathbb{P}\left(z_{\alpha/2} \leq \frac{\overline{X} - \lambda}{\sqrt{\lambda/n}} \leq z_{1-\alpha/2}\right) = \mathbb{P}\left(\left(\frac{\overline{X} - \lambda}{\sqrt{\lambda/n}}\right)^2 \leq z_{1-\alpha/2}^2\right) \approx 1 - \alpha.$$

Solving for λ in the middle expression gives the Wilson CI. We used one fewer approximation.

2 Confidence Intervals

2.1 Constructing CI

- Exact (i.e., coverage probability is exactly $1 - \alpha$) confidence intervals: **pivot method**:

  1. Find statistic $T_n = T(X, \theta)$ whose distribution is known and independent of θ. Such a statistic is called a pivot.

  2. Using knowledge on the distribution, find c_L and c_U such that $\mathbb{P}(c_L \leq T_n \leq c_R) = 1 - \alpha$.

  3. Convert the probability statement of the pivot to a probability statement of the parameter.

- Approximation methods, or asymptotic (sample size n is large) methods.

  - **Wald confidence interval** (the default CI produced by standard software): more than one approximations.

  - **Wilson's confidence interval** / score method / duality method: using CLT once.

    * Wilson's CI is better in the sense that the actual coverage is closer to the desired coverage.

  - **Variance stabilization transformation** (VST) method.

Example 2.1 (Pivot, Normal). Assume $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, $i = 1, \ldots, n$. Goal: $(1 - \alpha)$ CI for μ.

1. Recall that
$$T_n = \frac{\overline{X} - \mu}{\sqrt{S^2/n}} \sim t_{n-1}.[1]$$

2. We have
$$\mathbb{P}\left(t_{n-1,\alpha/2} \leq \frac{\overline{X} - \mu}{\sqrt{S^2/n}} \leq t_{n-1,1-\alpha/2}\right) = 1 - \alpha.$$

3. Rearrangement gives
$$\mathbb{P}\left(\overline{X} - t_{n-1,1-\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{n-1,1-\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

Example 2.2 (Wald, Bernoulli). Let $X_i \overset{\text{iid}}{\sim}$ Bernoulli(p), so that $\sum X_i \sim$ Binomial(n, p). Goal: $(1 - \alpha)$ CI for p. Note that $\hat{p} = \overline{X}$ is the MLE of p. CLT gives

$$\frac{\hat{p} - p}{\sqrt{\text{Var}(\hat{p})}} \approx \mathcal{N}(0, 1).$$

---

[1]t distribution has fatter tails than $\mathcal{N}(0, 1)$.

From Slutsky's theorem we have a **second approximation**

$$\text{Var}(\hat{p}) = \frac{p(1-p)}{n} \approx \frac{\hat{p}(1-\hat{p})}{n}.$$

This second approximation is characteristic of Wald's method. Using this second approximation we have

$$\frac{\overline{X} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} \approx \mathcal{N}(0,1),$$

with which we can construct the desired CI:

$$\left[\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\hat{p}(1-\hat{p})/n}, \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\hat{p}(1-\hat{p})/n} \right].$$

Example 2.3 (Wilson's CI; Bernoulli). We assume the same Bernoulli setup. Note that we obtained using just the CLT that

$$\mathbb{P}\left(z_{\frac{\alpha}{2}} \leq \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \leq z_{1-\frac{\alpha}{2}} \right) \approx 1 - \alpha.$$

We rewrite the middle expression as

$$n\,(\hat{p} - p)^2 \leq z_{1-\frac{\alpha}{2}}^2 \, p(1-p).$$

Solving a quadratic equation gives the desired Wilson's CI.

Remark 2.4.

- Note that the Wald CI is centered at $\hat{p}$, but Wilson's is not.

- Wilson's CI will always be contained in $[0, 1]$; the lower bound of Wald might be negative.

2.2 Variance Stabilization Transformation (VST) Method

Example 2.5 (VST, Poisson). Let $X_i \overset{\text{iid}}{\sim} \text{Poisson}(\lambda)$. Note that

$$\frac{\hat{\lambda} - \lambda}{\sqrt{\lambda/n}} \approx \mathcal{N}(0,1).$$

and

$$\sqrt{n}(\hat{\lambda} - \lambda) \approx \mathcal{N}(0, \lambda).$$

Goal: find a transformation (usually smooth) g such that

$$\sqrt{n}(g(\hat{\lambda}) - g(\lambda)) \approx \mathcal{N}(0, 1).[2]$$

---

[2]Or just a normal distribution with fixed variance.

Tool: delta method (Taylor expansion for random variables). By Taylor expansion,

$$g(\hat{\lambda}) = g(\lambda) + g'(\lambda)(\hat{\lambda} - \lambda) + \frac{g''(\lambda)}{2}(\hat{\lambda} - \lambda)^2 + \ldots.$$

We thus have the approximation

$$g(\hat{\lambda}) \approx g(\lambda) + g'(\lambda)(\hat{\lambda} - \lambda) + O(n^{-1}),$$

where the last term follows from the fact that $(\hat{\lambda} - \lambda)/\sqrt{\lambda/n} \approx \mathcal{N}(0, 1)$. Then,

$$\mathrm{E}[g(\hat{\lambda})] \approx g(\lambda) + g'(\lambda)\,\mathrm{E}[\hat{\lambda} - \lambda].$$

Since $\hat{\lambda}$ is unbiased, we have

$$\sqrt{n}(g(\hat{\lambda}) - g(\lambda)) \approx g'(\lambda)\sqrt{n}(\hat{\lambda} - \lambda),$$

giving

$$\sqrt{n}(g(\hat{\lambda}) - g(\lambda)) \approx \mathcal{N}(0, [g'(\lambda)]^2 \lambda),$$

where we used the approximation

$$\mathrm{Var}(g(\hat{\lambda})) = \mathrm{E}[(g(\hat{\lambda}) - \mathrm{E}[g(\hat{\lambda})])^2] \approx [g'(\lambda)]^2\,\mathrm{Var}(\hat{\lambda}).$$

To obtain $\sqrt{n}(g(\hat{\lambda}) - \lambda) \approx \mathcal{N}(0, 1)$, we need only $g'(\lambda)^2 = 1/\lambda$. $g(\lambda) = 2\sqrt{\lambda}$ will do.

That is, we have

$$\sqrt{n}(2\sqrt{\hat{\lambda}} - 2\sqrt{\lambda}) \approx \mathcal{N}(0, 1).$$

Using this we can obtain a CI for λ. Note that the left endpoint of the CI for $\sqrt{\lambda}$ may be negative, so in such cases when obtaining the CI for λ we need to use 0 instead.

2.3   Delta Method

Consider an unbiased estimator $\hat{\theta}$. $\mathrm{E}[\hat{\theta}] = \theta$. Assume that the variance is a function of the mean, $\mathrm{Var}[\hat{\theta}] = v(\theta)$.[3] Goal: find a function $g(t)$ such that $\mathrm{Var}[g(\hat{\theta})]$ is constant. Taylor gives

$$g(\hat{\theta}) \approx g(\theta) + g'(\theta)(\hat{\theta} - \theta).$$

Note that we have

$$\mathrm{E}[g(\hat{\theta})] \approx g(\theta) + g'(\theta)\,\mathrm{E}[\hat{\theta} - \theta] = g(\theta).$$

and

$$\mathrm{Var}[g(\hat{\theta})] = \mathrm{E}\left[\left(g(\hat{\theta}) - g(\theta)\right)^2\right] \approx \mathrm{E}\left[g'(\theta)^2\left(\hat{\theta} - \theta\right)^2\right] = g'(\theta)^2\,\mathrm{Var}[\hat{\theta}],$$

which we want to be constant, say 1. Thus we seek g such that

$$g'(\theta) = \frac{1}{\sqrt{v(\theta)}}.$$

A choice is of course $g(\theta) = \int 1/\sqrt{v(\theta)}\,d\theta$.

---

[3]This is a common case.

Example 2.6 (Delta Method, Exponential, distribution of MLE). Consider $X_i \overset{\text{iid}}{\sim} \text{Exponential}(\lambda)$. Recall that the MLE of λ is $\hat{\lambda} = 1/\overline{X}$. By the CLT we have for large n that

$$\frac{\overline{X} - \text{E}\,X}{\sqrt{\text{Var}\,\overline{X}}} \approx \mathcal{N}(0, 1),$$

That is,

$$\frac{\overline{X} - \frac{1}{\lambda}}{\sqrt{\frac{1}{n\lambda^2}}} = \sqrt{n}\left(\overline{X} - \frac{1}{\lambda}\right) \approx \mathcal{N}(0, \lambda^{-2}).$$

Goal: find an approximate distribution of $\hat{\lambda} = 1/\overline{X}$ using Delta Method. Using $g(t) = t^{-1}$ we have

$$\sqrt{n}\left(g(\hat{\lambda}) - g(\lambda)\right) \approx \mathcal{N}(0, [g'(\lambda)]^2 \,\text{Var}(\hat{\lambda})) = \mathcal{N}(0, \lambda^2).$$

(Note that Fisher's theorem gives another way to derive the above.)

Example 2.7 (VST, Exponential, CI). Consider the same setup as above. We can then derive an approximate CI for λ using VST. Recall that from Delta Method, we have

$$\sqrt{n}\left(g(\hat{\lambda}) - g(\lambda)\right) \approx \mathcal{N}(0, [g'(\lambda)]^2 \,\text{Var}(\hat{\lambda})).$$

Thus we seek g such that $g'(\lambda)^2\lambda^2$ is a constant, say 1, or $g'(\lambda) = 1/\lambda$. It is easy to see that $g = \log$ is such an option. Then,

$$\sqrt{n}\left(\log\hat{\lambda} - \log\lambda\right) \approx \mathcal{N}(0, 1),$$

using which we can obtain an approximate CI for $\log\hat{\lambda}$ and then $\hat{\lambda}$:

$$\left[\hat{\lambda}\exp\left(-\frac{z_{1-\alpha/2}}{\sqrt{n}}\right), \hat{\lambda}\exp\left(\frac{z_{1-\alpha/2}}{\sqrt{n}}\right)\right]$$

Theorem 2.8 (Delta Method). *If X_n is such that*

$$\sqrt{n}(X_n - \theta) \overset{\mathscr{D}}{\longrightarrow} \mathcal{N}(0, \sigma^2)$$

and g is continuously differentiable, then

$$\sqrt{n}(g(X_n) - g(\theta)) \overset{\mathscr{D}}{\longrightarrow} \mathcal{N}(0, \sigma^2[g'(\theta)]^2).$$

Remark 2.9. Intuition: We can write

$$\sqrt{n}(g(X_n) - g(\theta)) = g'(\tilde{\theta}_n)\sqrt{n}(X_n - \theta), \quad \tilde{\theta}_n \in (x_n, \theta).$$

We know that $g'(\tilde{\theta}_n) \to_{\text{p}} g'(\theta)$ and $\sqrt{n}(X_n - \theta) \overset{\mathscr{D}}{\to} \mathcal{N}(0, \sigma^2)$, so Slutsky's gives the desired result.

3 Change of Variable / Variable Transformation

3.1 The Discrete Case

Let X and $Y = g(X)$ be discrete. Then,

$$\mathbb{P}(Y = k) = \sum_{X:g(X)=k} \mathbb{P}(X = k).$$

3.2 The Continuous, 1D Case

Consider a continuous random variable X and transformation $Y = g(X)$ with g smooth with derivative vanishing nowhere, that is, $g' \neq 0$. These conditions guarantee that Y is also continuous. Suppose further that g is smooth and $g' \neq 0$. We have then that

$$f_Y(y) = f_X(x)\frac{1}{|g'(x)|} = f_X\left(g^{-1}(y)\right)\left|\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)\right|.$$

Example 3.1. Let $g' > 0$. Note that

$$\mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y)).$$

Thus

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}\mathbb{P}(Y \leq y) = \frac{\mathrm{d}}{\mathrm{d}y}F_X\left(g^{-1}(y)\right) = f_X\left(g^{-1}(y)\right)\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y).$$

Similarly, if $g' < 0$, then

$$\mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)).$$

Then

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}\mathbb{P}(Y \leq y) = -f_X(g^{-1}(y))\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y) = f_X(g^{-1}(y))\left|\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)\right|.$$

In general, if g is piecewise monotone, then

$$f_Y(y) = \sum_{\{x:g(x)=y\}} f_X(x)\frac{1}{|g'(x)|}.$$

Example 3.2. $Y = X^2$.

$$f_Y(y) = \sum_{\{x:x^2=y\}} f_X(x)\frac{1}{|2x|} = f_X(\sqrt{y})\frac{1}{2\sqrt{y}} + f_X(-\sqrt{y})\frac{1}{\left|-2\sqrt{y}\right|}.$$

3.3 The Continuous, 2D Case

Suppose X and Y are continuous with densities f_X and f_Y and joint density $f_{X,Y}$. Consider

$$\begin{cases} U = g_1(X, Y) \\ V = g_2(X, Y) \end{cases}$$

Suppose g_i's are smooth. Question: what is the joint density of U and V?

$$f_{U,V}(u, v) = f_{X,Y}(x, y)|J|,$$

where

$$J = \det \frac{\partial(x, y)}{\partial(u, v)} = \det \begin{bmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{bmatrix}$$

$$= \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

and we assume $J \neq 0$.

We have also

$$f_{XY}(x, y) = f_{UV}(u, v)|J|^{-1},$$

where

$$J^{-1} = \det \frac{\partial(u, v)}{\partial(x, y)} = \frac{1}{J}.$$

Example 3.3 (2D Change of Variables, Quotient). Let $U = X/Y$ with $Y \neq 0$ and $V = Y$. We have then that

$$\begin{cases} X = UV \\ Y = V \end{cases} .$$

Then

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{bmatrix} v & u \\ 0 & 1 \end{bmatrix}$$

so $J = v$. Alternatively,

$$\det \frac{\partial(u, v)}{\partial(x, y)} = \det \begin{bmatrix} \frac{1}{y} & -\frac{x}{y^2} \\ 0 & 1 \end{bmatrix} = \frac{1}{y} = J^{-1}.$$

This gives

$$f_{UV}(u, v) = f_{XY}(uv, v)|v|.$$

If we assume further that $X \perp\!\!\!\perp Y$, then

$$f_{UV}(u, v) = f_X(uv) f_Y(v)|v|.$$

11

4 Binomial and Poisson Distributions

Consider Binomial(n, p_n) with $np_n \to \lambda$ (thus $p_n \to 0$). We have then that

$$\mathbb{P}(X_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \frac{n!}{k!(n-k)!} \left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!} \frac{n}{n} \frac{n-1}{n} \cdots \frac{n-k+1}{n} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}.$$

Since k is fixed, as $n \to \infty$, and recalling that $(1 + X/n)^n \to e^X$, we have

$$\mathbb{P}(X = k) \to \frac{\lambda^k}{k!} e^{-\lambda}.$$

That is, if $np \to \lambda$, we have Binomial$(n, p) \to$ Poisson(λ). In this sense, the Poisson distribution is the limit of the binomial distribution, and we can thus use it as an approximation for the binomial distribution.

But note that for large n and np, since a Binomial is the sum of Bernoulli's, the CLT can be used to approximate the binomial distribution as $\mathcal{N}(np, np(1-p))$. Similarly, the Poisson distribution can be approximated as $\mathcal{N}(\lambda, \lambda)$.

Example 4.1. Let $X \sim$ Binomial(n, p). The CLT gives

$$\mathbb{P}\left(\frac{X - np}{\sqrt{np(1-p)}} \leq x\right) \approx \Phi(x)$$

and thus

$$\mathbb{P}(X = k) = \mathbb{P}(k - 1 < X \leq k) = \mathbb{P}\left(\frac{k-1-np}{\sqrt{np(1-p)}} < \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{k-np}{\sqrt{np(1-p)}}\right)$$

$$\approx \Phi\left(\frac{k-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-1-np}{\sqrt{np(1-p)}}\right).$$

It turns out that the **continuity correction** almost always gives a better approximation:

$$\mathbb{P}(X = k) = \mathbb{P}(k - 0.5 < X \leq k + 0.5) = \mathbb{P}\left(\frac{k-0.5-np}{\sqrt{np(1-p)}} < \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{k+0.5-np}{\sqrt{np(1-p)}}\right)$$

$$\approx \Phi\left(\frac{k+0.5-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-0.5-np}{\sqrt{np(1-p)}}\right).$$

5 Correlation

5.1 Bivariate Normal

Let (X, Y) be bivariate normal. That is, X and Y have joint density

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right).$$

If $X \perp\!\!\!\perp Y$, then $\rho = 0$ and the joint density simplifies to $f_X(x)f_Y(y)$.

5.2 Bivariate Normal, the Standard Case

The standard bivariate normal (U, V) is

$$f_{UV}(u, v) = \frac{1}{2\pi(1-\rho^2)} \exp\left(-\frac{1}{2(1-\rho^2)}(u^2 + v^2 - 2\rho uv)\right).$$

Note that

$$u^2 + v^2 - 2\pi uv = (v - \rho u)^2 + (1-\rho^2)u^2.$$

Thus we have

$$f_V(v) = \int_{-\infty}^{\infty} f_{UV}(u, v)\, dv = \frac{1}{2\pi(1-\rho^2)} \int_{-\infty}^{\infty} \exp\left(-\frac{(1-\rho^2)u^2}{2(1-\rho^2)}\right) \exp\left(-\frac{(v-\rho v)^2}{2(1-\rho^2)}\right)\, dv$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(v-\rho u)^2}{2(1-\rho^2)}\right)\, dv.$$

So the marginal distributions are also normal. One can verity similarly that $\mathrm{E}[UV] = \rho$, giving

$$\mathrm{Cov}(U, V) = \mathrm{E}(UV) - \mathrm{E}(U)\,\mathrm{E}(V) = \rho$$

and

$$\mathrm{Corr}(U, V) = \frac{\mathrm{Cov}(U, V)}{\mathrm{Var}(U)\,\mathrm{Var}(V)} = \rho.$$

Note that

$$f_{U|V}(u|v) = \frac{f_{UV}(u, v)}{f_V(v)} = \frac{\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(u^2 + v^2 - 2\rho uv)\right)}{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{v^2}{2}\right)}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(u^2 + v^2 - 2\pi uv - (1-\rho^2)v^2)\right)$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left(-\frac{(u-\rho v)^2}{2(1-\rho^2)}\right).$$

Thus,

$$U|V = v \sim \mathcal{N}(\rho v, (1-\rho^2))$$

and similarly
$$V|U = u \sim \mathcal{N}(\rho u, (1 - \rho^2)).$$

Thus the conditional expectation functions are
$$U = \rho V$$

and
$$V = \rho U.$$

Remark 5.1. This demonstrates regression toward the mean, since $\rho \leq 1$.

5.3 Bivariate Normal, the General Case

We may write
$$\begin{cases} X = \mu_X + \sigma_X U \sim \mathcal{N}(\mu_x, \sigma_X^2) \\ Y = \mu_Y + \sigma_Y V \sim \mathcal{N}(\mu_Y, \sigma_Y^2) \end{cases}.$$

We have
$$\mathrm{Cov}(X, Y) = \mathrm{Cov}(\mu_X + \sigma_X U, \mu_Y + \sigma_Y V) = \sigma_X \sigma_Y \, \mathrm{Cov}(U, V)$$

and thus
$$\mathrm{Corr}(X, Y) = \rho.$$

From
$$\frac{X - \mu_X}{\sigma_X} \bigg| \frac{Y - \mu_Y}{\sigma_Y} \sim \mathcal{N}\left(\rho \cdot \frac{y - \mu_Y}{\sigma_Y}, 1 - \rho^2\right)$$

we know
$$X|Y = y \sim \mathcal{N}\left(\mu_X + \rho \frac{\sigma_X}{\sigma_Y}(y - \mu_Y), \sigma_X^2(1 - \rho^2)\right).$$

Similarly,
$$Y|X = x \sim \mathcal{N}\left(\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1 - \rho^2)\right).$$

Note that the variance is smaller.

5.4 Bivariate Normal Data

Suppose (x_i, y_i) are iid bivariate normal. We have

$$\begin{aligned} L(\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho) &= \prod f(x_i, y_i) \\ &= \frac{1}{(2\pi)^2(1 - \rho^2)^{n/2}} \\ &\quad \exp\left(\frac{1}{2(1 - \rho^2)} \sum \left(\left(\frac{x_i - \mu_X}{\sigma_X}\right)^2 + \left(\frac{x_i - \mu_X}{\sigma_X}\right)^2 - 2\pi \left(\frac{x_i - \mu_X}{\sigma_X}\right)\left(\frac{y_i - \mu_Y}{\sigma_Y}\right)\right)\right). \end{aligned}$$

14

We have

$$\hat{\mu}_X = \overline{X}, \quad \hat{\mu}_X = \overline{Y}, \quad \hat{\sigma}_X = n^{-1}\sum(x_i - \overline{X})^2, \quad \hat{\sigma}_Y = n^{-1}\sum(y_i - \overline{Y})^2$$

and

$$\hat{\rho} = \frac{\sum(x_i - \overline{X})(y_i - \overline{Y})}{\sqrt{\sum(x_i - \overline{X})^2 \sum(y_i - \overline{Y})^2}}.$$

We will use VST to get the Fisher transformation

$$\frac{1}{2}\log\frac{1+\rho}{1-\rho}.$$