Bivariate Distributions

Independent: $f(x) \cdot f(y) = f(x,y)$

$E[Y_1] = \int_0^1 y_1 \cdot$ (Marginal)

Example: $f(x,y) = k(x+2y)$
where $0 < 2y < x < 2$

$k \cdot \int_0^2 \int_0^{x/2} x+2y \, dy dx$

$k \cdot \int_0^1 \int_{2y}^2 x+2y \, dx dy$

Area under: Inner $\int_f^1$
Area over: Inner $\int_0^f$
} Look where the „other" function enters and leaves the area and which value it is.

CDF

$F(y) = \int_0^y \dots dy$

Is CDF, when
① $F(-\infty) = \lim_{y \to \infty} = 0$
② $F(\infty) = \lim_{y \to \infty} = 1$
③ $F(y)$ is monotonically increasing

Density Function

$P(a \le y \le b) = \int_a^b f(y) dy$

$= F(b) - F(a)$

Properties of Variance

$V[c] = 0$

$V[Y+c] = V[c]$

$V[c \cdot y] = c^2 \cdot V[y]$

Partial Integration

$\int u' \cdot v = u \cdot v - \int u \cdot v'$

Select v that v' gets easier.

Mean and Median

Median: $\int_{-\infty}^{y} f(y) dy = 0.5$

Mean: $E[Y] = \int_{-\infty}^{\infty} y \cdot f(y) dy$

$E[g(y)] = \int_{-\infty}^{\infty} g(y) \cdot f(y) dy$

Unbiased

The mean of the estimate is equal to the parameter

Consistent

The variance of the estimator goes to 0 as $n \to \infty$.

V and E

Rules for E and V

E: ① constants out
② $E(\sum y) = \sum E(y)$

V: ① constants out + square
② $V(\sum y) = \sum V(y)$

$V[Y] = E[Y^2] - E^2[Y]$

Method of Moments

$\mu'_k = E[Y^k]$ Population Moments

$m'_k = 1/n \cdot \sum_{i=1}^{n} y_i^k$ Sample Moments

k = # parameters to estimate

Solve $\mu'_k = m'_k$

Maximum Likelyhood

$L(\lambda) = \prod_{i=1}^{n} f(y)$

Find maximum

Hypothesis Testing

$P(Z - \alpha/2 \le \frac{x-\mu}{\sigma} \le Z\,\alpha/2)$ $\begin{matrix}\mu = 10 \\ \sigma = 2\end{matrix} \Big| \begin{matrix}\cdot\sigma \\ +10\end{matrix}$

$N(10,4)$ $\downarrow$ mean 10, deviation 2 (square)

$\Downarrow$

$Z = \frac{x - \mu}{\sigma} \sim N(0,1)$

Marginals

Conditional Probability

$P(X|Y) = \frac{P(X,Y)}{P(y)}$

$f(y_1^x) = \int f(y_1, y_2) \, dy_2$ } limit around these, so x in limits

$f(y_2) = \int f(y_1, y_2) \, dy_1$

$E[X|Y=0.5] = \int P(x|y=0.5) \cdot \underline{x} \, dx$

$MSE = S^2 = \frac{SSE}{n-2}$ unbiased estimator of σ^2

Linear Regression

$E[Y] = \beta_0 + \beta_1 \cdot X$

$V[Y] = \sigma^2$

$\beta_1 = \frac{Sxy}{Sxx}$

$\beta_0 = \bar{y} - \bar{\beta_1} \cdot \bar{x}$

For SSE: calculate difference between function values and given ones. Square. Add. Profit.

Table

Y	X	$Y - \bar{Y}$	$X - \bar{X}$	$(X - \bar{X})^2$	$(X - \bar{X})(Y - \bar{Y})$
				$\downarrow$	$\downarrow$
				$\sum Sxx$	Sxy
				$\downarrow$	$\downarrow$
				$\sum_{i=1}^{n} (x_i - \bar{x})^2$	$\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$

$SSE = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = e_i^2$

$\beta_0: T = \frac{\hat{\beta}_0 - 0}{S \cdot \sqrt{\frac{\sum x_i^2}{n \cdot Sxx}}}$ ← Testvalue; → Original x values, square, sum, Profit

„check if T is inside region"

$\beta_1: T = \frac{\hat{\beta}_1 - 0}{\sqrt{\frac{\sigma^2}{Sxx}}}$ $= \sigma^2 = \frac{SSE}{n-2}$ n = # parameters

$\sigma = S$ T = Acceptance Reg

Suggest estimator

$\bar{y} = 1/n \sum_{i=1}^{n} y_i$ check $E(y) = \lambda$

$cov(y_1, y_2) = E[y_1 \cdot y_2] - \mu_1 \mu_2$

$cov(y_1, y_2) = 0$ if Y_1, Y_2 are independent

if $Y_1 \perp Y_2$ then $cov(Y_1, Y_2) = E[Y_1] \cdot E[Y_2] = \mu_1 \cdot \mu_2$

$U = \sum_{i=1}^{n} a_i Y_i$ and $V = \sum_{j=1}^{m} b_j X_j$, then

a) $E[U] = \sum_i a_i \mu_i$ b) $V[U] = \sum_i a_i^2 \cdot Var(Y_1) +$

$2 \cdot \sum \sum_{1 \le i \le j \le n} a_i a_j \, cov(y_i, y_j)$

c) $cov(U,V) = \sum_i \sum_j a_i b_j \, cov(Y_i, X_j)$

Sampling Distribution

$\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i$

$\chi^2(n) \sim \frac{(n-1) \cdot S^2}{\sigma^2}$ (n ← df)

$S^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2$

$t(v) \sim T = \frac{Z}{\sqrt{\frac{W}{v}}}$ $Z = N(0,1)$ $Z \perp W$

Likelyhood

$Z = \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}}$ } pivotal quantities

$\hat{\theta} \pm Z_{\alpha/2} \sigma_{\hat{\theta}}$

Sum + Product

$\sum_{i=1}^{n} c = n \cdot c$ $\prod_{i=1}^{n} c = c^n$

$\chi^2(n) = \frac{(n-1) S^2}{\sigma^2}$

Population variance ← σ^2; S^2 → sample variance

n = df

Bayesian Method for y estimate

① Prior $P(\theta)$. Replace λ with θ

② Likelyhood $P(y|\theta)$ is $L(\lambda)$ → from LME

③ Write $P(\theta|y) \propto \underbrace{P(\theta) \cdot P(y|\theta)}_{\text{ignore constants}}$ Posterior

④ Form to known distribution
Conjugate says, which family

⑤ Get α^x (and β^x), spot difference

⑥ Probably E of distribution with new α^x and b^x

⑦ For k substitute $\alpha + \beta$ in the dropped out constant with $\alpha^x + \beta^x$. Other values should be given.

Confidence Intervals

$100(1-\alpha)\%$

CI for $E[y] = \beta_0 + \beta_1 X^x$

$\hat{\beta}_i \pm t_{\alpha/2} \cdot S \cdot \sqrt{c_{ii}}$

$c_{00} = \frac{\sum x_i^2}{n \cdot Sxx}$ $c_{11} = \frac{1}{Sxx}$

↳ is:

$\hat{\beta}_0 + \hat{\beta}_1 X^x \pm t_{\alpha/2} \cdot t(n-2) \cdot \frac{1}{n} \cdot \sqrt{\frac{(X^x - \bar{x})^2}{Sxx}} \cdot S$

Prediction Intervals

$100(1-\alpha)\%$ PI for y when $X = x^x$

$\hat{\beta}_0 + \hat{\beta}_1 X^x \pm t_{\alpha/2} \cdot S \sqrt{1 + 1/n + \frac{(X^x - \bar{x})^2}{Sxx}}$

σ^2 ← ; → S^2

Point Estimators

target parameter θ	sample size	point estimator $\hat{\theta}$	$E(\hat{\theta})$	std error $\sigma_{\hat{\theta}}$
μ	n	$\bar{Y}$	μ	$\sigma/\sqrt{n}$
p	n	$\hat{p} = Y/n$	p	$\sqrt{p \cdot q/n}$
$\mu_1 - \mu_2$	n_1, n_2	$\bar{Y}_1 - \bar{Y}_2$	$\mu_1 - \mu_2$	$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$p_1 - p_2$	n_1, n_2	$\hat{p}_1 - \hat{p}_2$	$p_1 - p_2$	$\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$

Properties of Least-Squares Estimators Simple Linear Regression

① $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased, $E[\hat{\beta}_i] = \hat{\beta}$

② $V(\hat{\beta}_0) = \sum(x_i^2/(n \cdot S_{xx})) \cdot \sigma^2$

③ $V(\hat{\beta}_1) = \sigma^2/S_{xx}$

④ $cor(\hat{\beta}_0, \hat{\beta}_1) = \frac{-\bar{x}}{S_{xx}} \cdot \sigma^2$

⑤ Unbiased estimator of σ^2 is $S^2 = SSE/(n-2)$ where $SSE = S_{yy} - \hat{\beta}_1 S_{xy}$ and $S_{yy} = \sum(y_i - \bar{y})^2$

If ε_i are normally distributed

⑥ $\hat{\beta}_0, \hat{\beta}_1$ are normally distributed

⑦ $\frac{(n-2)S^2}{\sigma^2}$ has χ^2 with n-2 df

⑧ Statistics S is independant of both $\hat{\beta}_0$ and $\hat{\beta}_1$

$\perp$ = independent

General Linear Models

OLS solution:

$\hat{\beta} = (X^T X)^{-1} X^T Y$

$SSE = Y^T Y - \hat{\beta} X^T Y$

$Y^T Y = \sum Y_i^2$

$S^2 = \frac{SSE}{n-2}$

t.ex. Test H_0: $a^T\beta = (a^T\beta)_0$

$$T = \frac{a^T\beta - a^T\beta_0}{S \cdot \sqrt{a^T (X^T X)^{-1} a}}$$

A $100(1-\alpha)\%$ CI for $a^T\beta$ is given by

$$a^T\beta \pm \underline{t_{\alpha/2}} S \sqrt{a^T (X^T X)^{-1} a}$$

↳ $t(n-k-1)$ df

k: order of the moment, #parameters -1

k: highest $\beta_x \rightarrow x = k$

t.ex inference for β_1? $a = (0,1,0,\dots)$

$a^T\beta = \beta_1$

Bayes Estimator

$\hat{\theta} = E[p(\theta|y)]$

↳ „Posterior Mean"

The posterior mean minimizes the MSE = $E[(\hat{\theta}-\theta)^2]$

MAA is without square

Gamma Function

$\Gamma(n) = (n-1)!$

$0! = 1! = 1$

x^* with $E[Y] =$

$$\beta_0 + \beta_1 \cdot x^* \pm t_{\alpha/2} \cdot S \cdot \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}$$

⇑ CI

↓ plus

PI: for Y when $x = x^*$

$$\beta_0 + \beta_1 \cdot x^* \pm t_{\alpha/2} \cdot S \cdot \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}}$$

$E[\bar{Y}] = E[1/n \cdot \sum Y] = 1/n \sum E[Y] = 1/n \cdot n \cdot \mu = \mu$

$V[\bar{Y}] = V[1/n \sum Y] = 1/n^2 \cdot \sum V[Y] = 1/n^2 \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n} \rightarrow 0$ as $n \rightarrow \infty$

↳ consistent

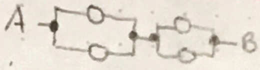
Poisson

$e^{-\theta(n + 1/\beta)} \Rightarrow \beta^* = \frac{1}{n + 1/\beta}$

H_0-Testing (for LR)

We have two approaches:

T-Testing or CI-Testing



$P = (0.9 + 0.9 - (0.9)^2) \cdot (0.9 + 0.9 - (0.9)^2)$

T-Testing

- Calculate T
- look up $t_{\alpha/2}$
- $|T| \geq t$ reject
- $|T| < t$ don't reject

$$T = \frac{a^T\beta - a^T\beta_0}{S \cdot \sqrt{a^T (X^T X)^{-1} a}}$$

CI-Testing

- Calculate CI with formula
- Check if real value lies within

$$a^T\beta \pm t_{\alpha/2} \cdot S \cdot \sqrt{a^T (X^T X)^{-1} a}$$