STAT 712 fa 2022 Lec 11 slides

Sundry bivariate nuggets, some inequalities, hierarchical models

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

$$\mathrm{Cov}(X,X) = \mathbb{E}(X-\mu_X)(X-\mu_X) = \mathbb{E}(X-\mu_X)^2 = \mathrm{Var}\,X$$

$$\mathrm{Var}\,X = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

Covariance

The covariance between two rvs X and Y is defined as

$$\mathrm{Cov}(X,Y) = \mathbb{E}(X-\mu_X)(Y-\mu_Y) =: \underline{\sigma_{XY}},$$

where $\mu_X = \mathbb{E}X$ and $\mu_Y = \mathbb{E}Y$.

Useful expression: $\mathrm{Cov}(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$

$\mathbb{E}XY - \mu_X\mu_Y$

Exercise: Derive the useful expression for computing covariances.

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

$$= \mathbb{E}XY - \mu_X\mu_Y - \mu_Y\mu_X + \mu_X\mu_Y =$$

Correlation

The correlation between two rvs X and Y is defined as

$$\operatorname{corr}(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var} X}\sqrt{\operatorname{Var} Y}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} =: \rho_{XY},$$

where $\sigma_X = \sqrt{\operatorname{Var} X}$ and $\sigma_Y = \sqrt{\operatorname{Var} Y}$.

Theorem (Correlation between minus 1 and 1, cf. Thm 4.5.7 in CB)

For any rvs X and Y,

1. $-1 \leq \operatorname{corr}(X, Y) \leq 1$
2. $\operatorname{corr}(X, Y) = \pm 1$ *iff there exist* $a \neq 0$ *and* b *such that* $P(Y = aX + b) = 1$.

We will prove the first part of this result later.

$$\mathrm{Var}\,X = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

Exercise: Let (X, Y) be a pair of rvs with joint pdf given by

$$f(x, y) = \frac{1}{8}(x + y) \cdot \mathbf{1}(0 < x < 2, 0 < y < 2).$$

$$f_X(x) = \int_0^2 \frac{1}{8}(x+y)\,dy = \frac{1}{8}\left(xy + \frac{y^2}{2}\right)\Bigg|_0^2 = \frac{1}{8}\left(2x + \frac{4}{2}\right) = \frac{x+1}{4} \quad 0<x<2.$$

1. Find $\mathrm{Cov}(X, Y)$. $= \mathbb{E}XY - \mathbb{E}X\,\mathbb{E}Y = \frac{4}{3} - \left(\frac{7}{6}\right)^2.$
2. Find $\mathrm{corr}(X, Y)$.

$$\mathbb{E}XY = \int_0^2\int_0^2 x\cdot y \; \frac{1}{8}(x+y)\,dx\,dy = \dots = \frac{4}{3}$$

$$\mathbb{E}X = \int_0^2 x \cdot \left(\frac{x+1}{4}\right) dx = \dots = \frac{7}{6}, \quad \mathbb{E}Y = \frac{7}{6}$$

$$\mathrm{corr}(X,Y) = -\frac{1}{11}.$$

$$\text{Var}(aX+b) = a^2 \text{Var} X \qquad \text{sign}(z) = \begin{cases} -1 & \text{if } z<0 \\ 0 & \text{if } z=0 \\ 1 & \text{if } z>0 \end{cases}$$

Covariance and correlation of linearly transformed rvs

For any two rvs X and Y and constants $a, b, c, d \in \mathbb{R}$, we have

$$\begin{aligned} \text{Cov}(aX+b, cY+d) &= ac \cdot \text{Cov}(X,Y) \\ \text{corr}(aX+b, cY+d) &= \text{sign}(ac) \cdot \text{corr}(X,Y). \end{aligned}$$

Exercise: Prove the result.

$$(a\,\mathbb{E}X+b)\cdot(c\cdot\mathbb{E}Y+d)$$

$$① \quad \text{Cov}(aX+b, cY+d) = \mathbb{E}\left[(aX+b)(cY+d)\right] - \mathbb{E}(aX+b)\cdot\mathbb{E}(cY+d)$$

$$= \mathbb{E}\left[acXY + \cancel{adX} + \cancel{bcY} + \cancel{bd}\right]$$

$$- \left[ac\,\mathbb{E}X\,\mathbb{E}Y + \cancel{ad\,\mathbb{E}X} + \cancel{bc\,\mathbb{E}Y} + \cancel{bd}\right]$$

$$= ac\underbrace{(\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y)}$$

$$= \underline{a \cdot c \; \mathrm{Cov}(X, Y)}$$

$$\mathrm{Cov}(aX+b, aX+b) = a \cdot a \cdot \mathrm{Cov}(X, X) = a^2 \, \mathrm{Var}\, X.$$

② $$\mathrm{Corr}(aX+b, cY+d) = \frac{a \cdot c \cdot \mathrm{Cov}(X, Y)}{\sqrt{\underline{\mathrm{Var}(aX+b)} \; \underline{\mathrm{Var}(cY+d)}}}$$

$$= \frac{a \cdot c \cdot \mathrm{Cov}(X, Y)}{\sqrt{a^2 \, \mathrm{Var}\, X \quad c^2 \, \mathrm{Var}\, Y}}$$

$\sqrt{a^2} = |a|$

$$= \frac{a}{|a|} \frac{c}{|c|} \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}\, X \; \mathrm{Var}\, Y}}$$

$\frac{a}{|a|}$

$$= \mathrm{sign}(a) \cdot \mathrm{sign}(c) \; \mathrm{Corr}(X, Y)$$

$$= \mathrm{sign}(a \cdot c) \; \mathrm{Corr}(X, Y).$$

$X \perp\!\!\!\perp Y \quad \substack{\Rightarrow \\ \not\Leftarrow} \quad \mathrm{Cov}(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$
$= \mathbb{E}X\mathbb{E}Y - \mathbb{E}X\mathbb{E}Y = 0$

Theorem (independence implies covariance equal to zero)

If X and Y are independent then $\mathrm{Cov}(X, Y) = 0$.

$f_X(x)$

$1/2$

$-1 \quad 0 \quad 1$

If $\mathrm{Cov}(X, Y) = 0$, it does not mean that X and Y are independent!

Exercise: Let (X, Y) be a pair of rvs with joint pdf given by

$$f(x, y) = \frac{1}{2|x|} e^{-y/|x|} \mathbf{1}(x \in (-1, 1) \setminus \{0\}, y > 0).$$

1. Check whether X and Y are independent.
2. Compute $\mathrm{Cov}(X, Y)$. $= \underbrace{\mathbb{E}XY}_{0} - \underbrace{\mathbb{E}X}_{0}\mathbb{E}Y = 0$

$X \not\perp\!\!\!\perp Y$

$f_X(x) = \int_0^\infty \frac{1}{2|x|} e^{-y/|x|} \, dy$

$= \frac{1}{2} \cdot 1 = \frac{1}{2}$ for $x \in (-1,1) \setminus \{0\}$

$\mathbb{E}XY = 0$

$\mathbb{E}X = 0$

$$\mathbb{E}XY = \int_{-1}^{1}\int_{0}^{\infty} x \cdot y \, \frac{1}{2|x|} e^{-y/|x|} \, dy \, dx$$

$$= \int_{-1}^{1} \frac{x}{2} \underbrace{\int_{0}^{\infty} y \cdot \frac{1}{|x|} e^{-y/|x|} \, dy}_{|x|} \, dx$$

$$= \int_{-1}^{1} \frac{x \cdot |x|}{2} \, dx$$

$$= \frac{1}{2}\int_{-1}^{0} x(-x) \, dx + \frac{1}{2}\int_{0}^{1} x \cdot x \, dx$$

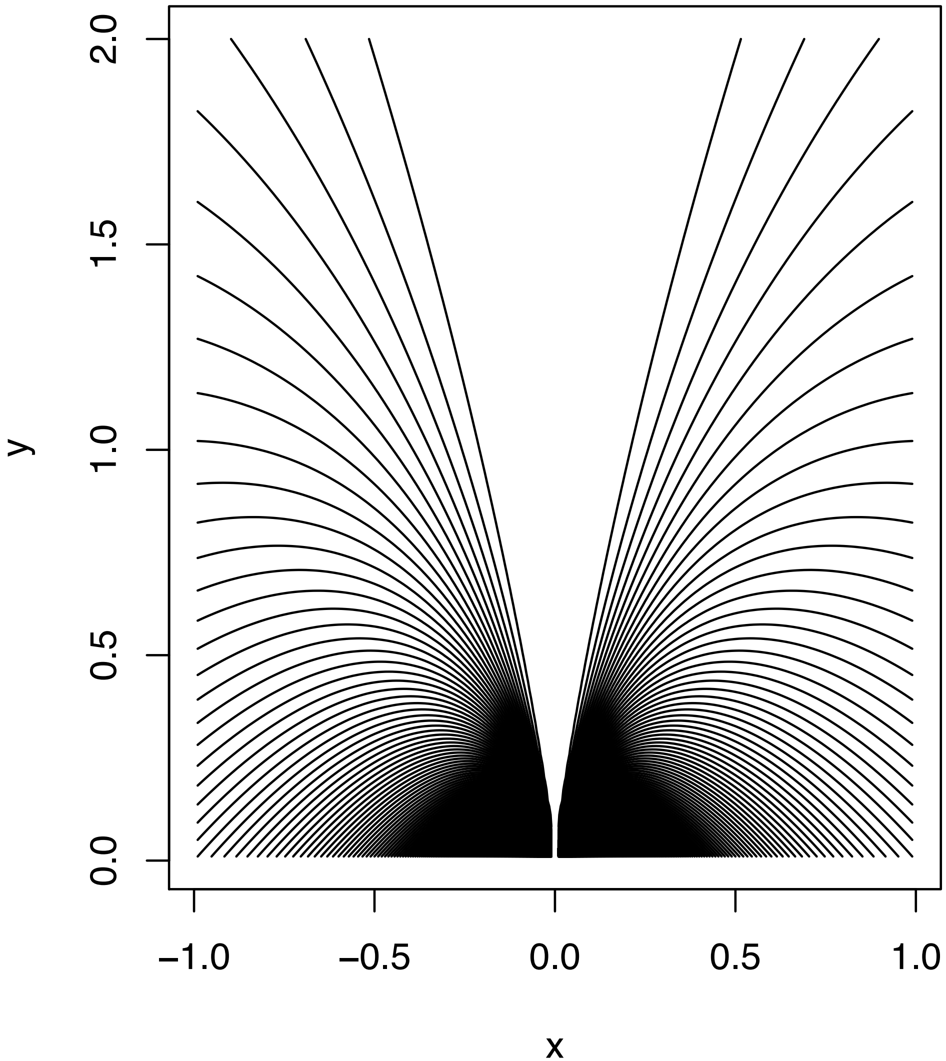
$$= -\frac{1}{2}\int_{-1}^{0} x^2 \, dx + \frac{1}{2}\int_{0}^{1} x^2 \, dx$$

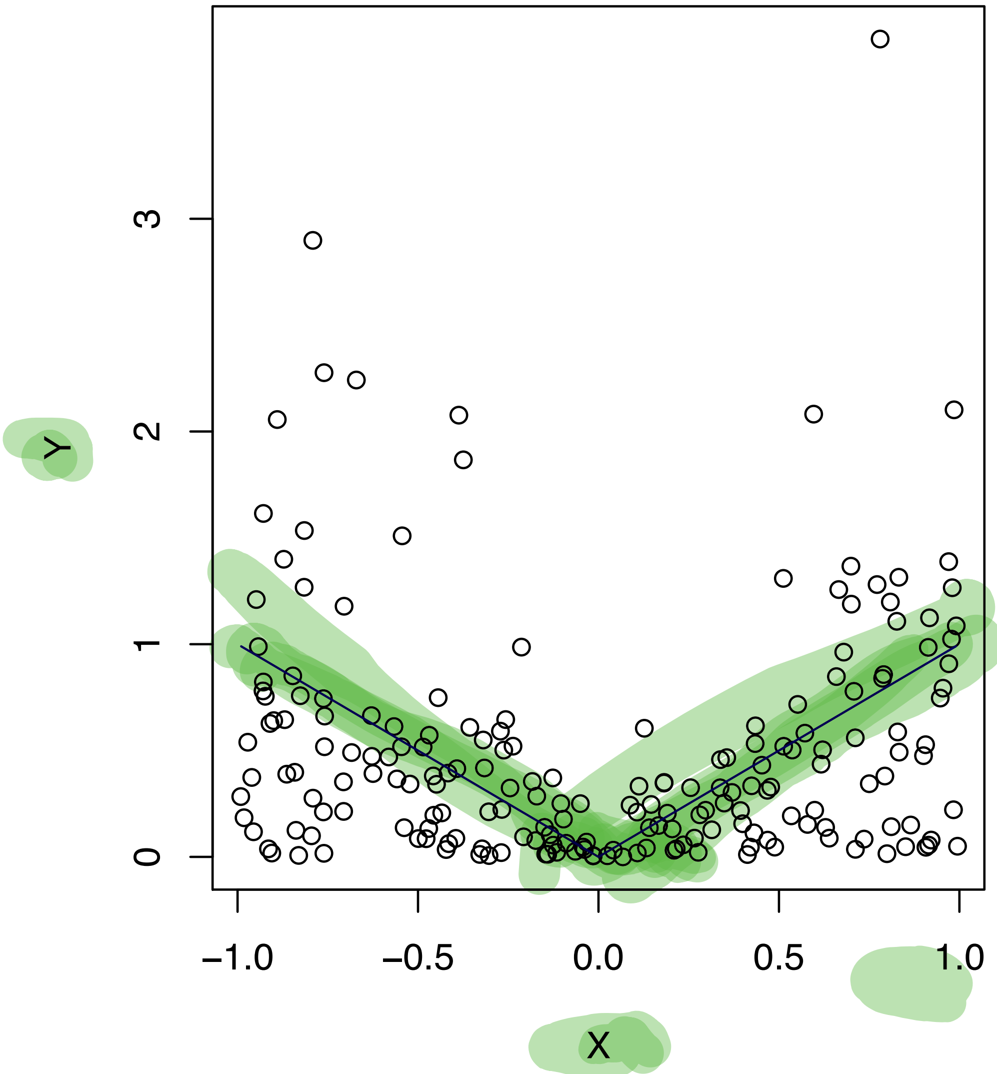
$$= -\frac{1}{2}\left[\frac{x^3}{3}\right]\Bigg|_{-1}^{0} + \frac{1}{2}\left[\frac{x^3}{3}\right]\Bigg|_{0}^{2}$$

$$= -\frac{1}{2}\left(0 - \left(\frac{-1}{3}\right)\right) + \frac{1}{2}\,\frac{1}{3}$$

$$= -\frac{1}{6} + \frac{1}{6}$$

$$= 0$$


y
x
0.0
0.5
1.0
1.5
2.0
−1.0
−0.5
0.0
0.5
1.0


Y
X
0
1
2
3
−1.0
−0.5
0.0
0.5
1.0

$\sum_{i=1}^{n} a_i X_i \qquad \sum_{j=1}^{m} b_j Y_j$

Theorem (Variance/Covariance of linear combinations of rvs)

Let $X_1, \dots, X_n$ *and* $Y_1, \dots, Y_m$ *be rvs and let* $a_1, \dots, a_n, b_1, \dots, b_m \in \mathbb{R}$. *Then*

1. $\text{Cov}(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j) = \underline{\sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(X_i, Y_j)}$.
2. $\text{Var}(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + 2\sum_{i<j} a_i a_j \text{Cov}(X_i, X_j)$.

Simple cases of the above are

1. $\text{Cov}(aX + bY, cU + dV) =$
$$ac \text{Cov}(X, U) + ad \text{Cov}(X, V) + ac \text{Cov}(Y, U) + bd \text{Cov}(Y, V).$$
2. $\text{Var}(aX + bY) = a^2 \text{Var} X + b^2 \text{Var} Y + 2ab \text{Cov}(X, Y)$.

Exercise: Prove the general results.

① $\mathrm{Cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right)$

$$= \mathbb{E}\left(\sum_{i=1}^{n} a_i X_i \sum_{j=1}^{m} b_j Y_j\right) - \mathbb{E}\left(\sum_{i=1}^{n} a_i X_i\right)\mathbb{E}\left(\sum_{j=1}^{m} b_j Y_j\right)$$

$$= \mathbb{E}\left(\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j X_i Y_j\right) - \sum_{i=1}^{n} a_i \mathbb{E}X_i \sum_{j=1}^{m} b_j \mathbb{E}Y_j$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \mathbb{E}X_i Y_j - \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \mathbb{E}X_i \mathbb{E}Y_j$$

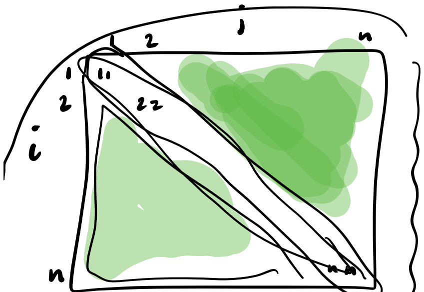
$$= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \left(\underbrace{\mathbb{E}X_i Y_j - \mathbb{E}X_i \mathbb{E}Y_j}\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \mathrm{Cov}(X_i, Y_j) = \underset{\sim}{a}^T \Sigma \underset{\sim}{b}$$

$$\underset{\sim}{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \quad \underset{\sim}{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix} \quad \underset{\sim}{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \quad \underset{\sim}{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

$$\underset{n\times m}{\Sigma} = \Big(\mathrm{Cov}(X_i, Y_j)\Big)_{1\le i\le n,\ 1\le j\le m}$$

② $\mathrm{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \mathrm{Cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{n} a_j X_j\right)$



$(i,j) = (2,3)$ $\cancel{(3,2)}$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{n} a_i^2 \mathrm{Var}(X_i) + \sum_{i\neq j} a_i a_j \mathrm{Cov}(X_i, X_j)$$

$$= \sum_{i=1}^{n} a_i^2 \mathrm{Var}(X_i) + 2\sum_{i<j} a_i a_j \mathrm{Cov}(X_i, X_j)$$

$$\mathrm{Corr}(Z_i, Z_j) = \frac{\mathrm{Cov}(Z_i, Z_j)}{\sqrt{\underbrace{\mathrm{Var}\,Z_i}_{=1}\ \underbrace{\mathrm{Var}\,Z_j}_{=1}}} = \mathrm{Cov}(Z_i, Z_j)$$

$\mathrm{Var}\,Z_i = 1$ for all i

Exercise: Let $Z_1, \ldots, Z_n$ have unit variance and suppose

$$\mathrm{corr}(Z_i, Z_j) = \rho \in (-1, 1) \text{ for } i \neq j.$$

Find $\mathrm{Var}\,\bar{Z}$, where $\boxed{\bar{Z} = n^{-1}\sum_{i=1}^n Z_i.}$

$$\mathrm{Var}(\bar{Z}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n Z_i\right)$$

$$= \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^n Z_i\right)$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^n \overbrace{\mathrm{Var}\,Z_i}^{1} + 2\sum_{i<j} \overbrace{\mathrm{Cov}(Z_i, Z_j)}^{\rho}\right]$$

$$= \frac{1}{n^2}\left[\underbrace{\sum_{i=1}^{n} 1}_{n} + 2\underbrace{\sum_{i<j}}_{\binom{n}{2}} \rho\right]$$

$$= \frac{1}{n^2}\left[n + 2\frac{n!}{(n-2)!\,2!}\rho\right]$$

$$= \frac{1}{n^2}\left[n + n(n-1)\rho\right]$$

$$\mathrm{Var}\,\bar{Z} = \frac{1}{n}\left[1 + (n-1)\rho\right]$$

Exercise: Let $Y_1, \ldots, Y_n$ be independent rvs such that

$$Y_i \sim \text{Normal}(\mu, \sigma_i^2), \quad i = 1, \ldots, n$$

and let

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \quad \text{and} \quad \tilde{Y} = \frac{\sum_{i=1}^{n} \sigma_i^{-2} Y_i}{\sum_{j=1}^{n} \sigma_j^{-2}}.$$

1. Find $\mathbb{E}\bar{Y}$.
2. Find $\text{Var}\,\bar{Y}$.
3. Find $\mathbb{E}\tilde{Y}$.
4. Find $\text{Var}\,\tilde{Y}$.
5. Consider case $\sigma_1^2 = \cdots = \sigma_n^2$.

Bivariate Normal distribution

The rvs (X, Y) have the bivariate Normal distribution if they have joint pdf

$$f(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho) = \frac{1}{2\pi} \frac{1}{\sigma_X \sigma_Y \sqrt{1-\rho^2}}$$

$$= \begin{vmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{vmatrix}^{1/2}$$

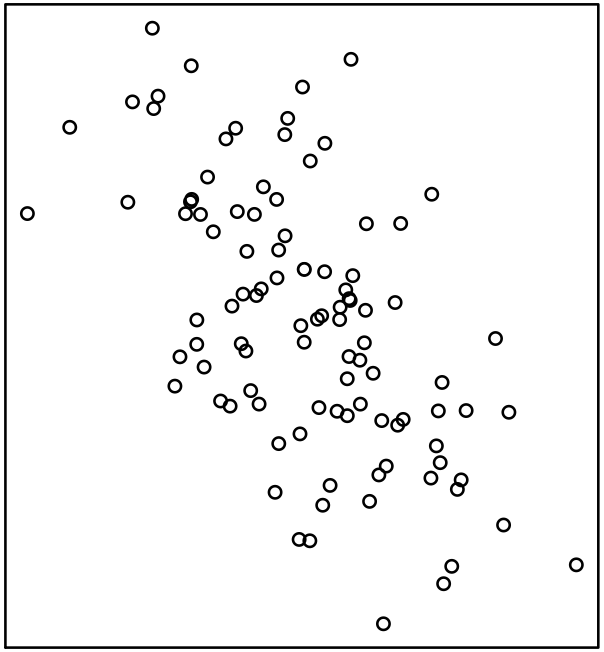
$$\times \exp\left[-\frac{1}{2(1-\rho^2)} \underbrace{\left(\left[\frac{X-\mu_X}{\sigma_X}\right] - 2\rho \left[\frac{X-\mu_X}{\sigma_X}\right]\left[\frac{Y-\mu_Y}{\sigma_Y}\right] + \left[\frac{Y-\mu_Y}{\sigma_Y}\right]^2 \right)} \right],$$

where

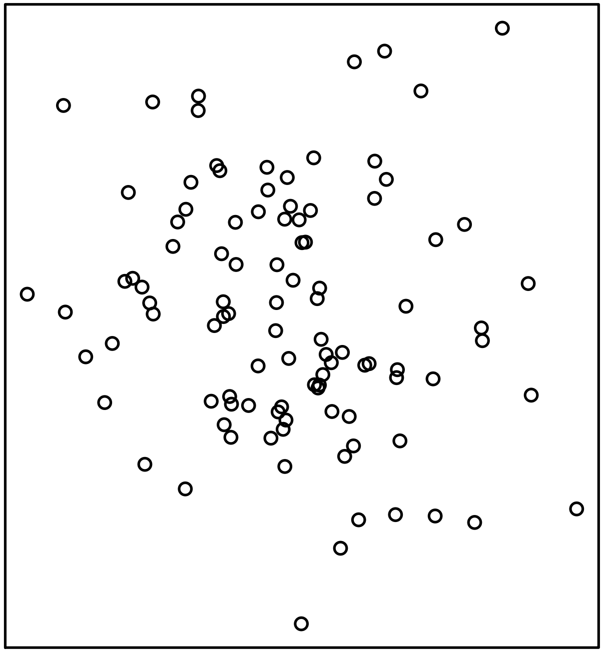
- μ_X and μ_Y are mean of X and Y.
- σ_X^2 and σ_Y^2 are variance of X and Y.
- ρ is $\text{corr}(X, Y)$.

$\rho = 0$

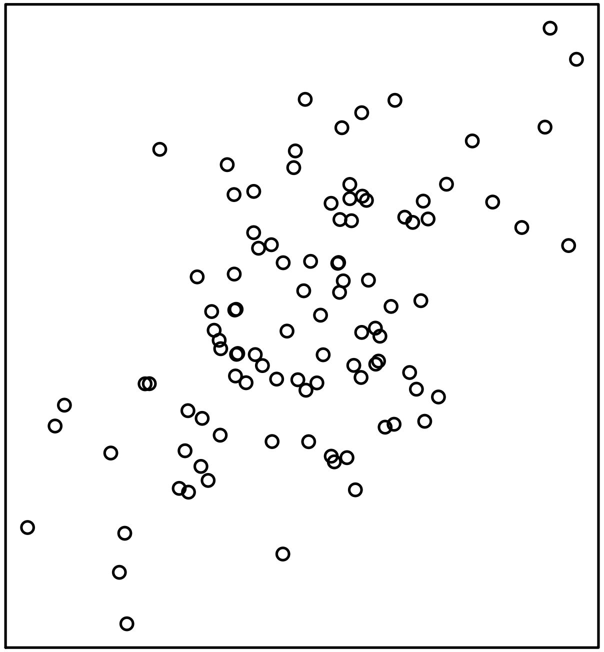
Exercise: Show biv-Normal (X, Y) are independent iff $\text{corr}(X, Y) = 0$.



rho = −0.5



rho = 0



rho = 0.5

Exercise: Show that with $\rho = \sigma_{XY}/(\sigma_X \sigma_Y)$, we can write the biv-Normal pdf as

$$f(x,y) = \frac{\exp\left[-\frac{1}{2}\begin{pmatrix} X-\mu_X \\ Y-\mu_Y \end{pmatrix}^T \begin{pmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{pmatrix}^{-1} \begin{pmatrix} X-\mu_X \\ Y-\mu_Y \end{pmatrix}\right]}{2\pi \cdot \begin{vmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{vmatrix}^{1/2}}.$$

└ Covariance matrix of (X, Y)

Multivariate Normal distribution

Let $\mathbf{X} = (X_1, \dots, X_d)^T$ be a vector of rvs with joint pdf

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}|^{1/2}} \exp\left[-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right],$$

where $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\mathbf{\Sigma}$ a symmetric positive definite $d \times d$ matrix.

Then we say $\mathbf{X}$ has the *multivariate Normal distribution* with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, where $\mathbb{E}X_j = \mu_j$ and $\text{Cov}(X_j, X_k) = \mathbf{\Sigma}_{jk}$, $1 \leq j, k \leq d$.

Theorem (Cauchy-Schwarz Inequality)

For any rvs X *and* Y *we have* $|\mathbb{E}XY| \leq \mathbb{E}|XY| \leq \sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}Y^2}$

Exercise:

1. Prove the inequality.
2. Use to prove $\text{corr}(X, Y) \in [-1, 1]$ for any rvs X, Y.

(a) Show $|\mathbb{E}XY| \leq \mathbb{E}|XY|$.

We have

$-a < x < a$

$\iff |x| \leq a$

$$-|XY| \leq XY \leq |XY|$$

Take expectations:

$$-\mathbb{E}|XY| \leq \mathbb{E}XY \leq \underbrace{\mathbb{E}|XY|}_{\geq 0}$$

$$\iff |\mathbb{E}XY| \leq \mathbb{E}|XY|$$

(b) Show $\mathbb{E}|XY| \leq \sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}Y^2}$.

To start: For any $a, b \in \mathbb{R}$,

$$0 \leq (a-b)^2 = a^2 + b^2 - 2ab$$

$$\iff a^2 + b^2 \geq 2ab$$

$$\iff \boxed{ab \leq \frac{a^2+b^2}{2}}$$

Now let $a = \dfrac{|X|}{\sqrt{\mathbb{E}X^2}}$, $b = \dfrac{|Y|}{\sqrt{\mathbb{E}Y^2}}$

Then $$ab = \frac{|X|\,|Y|}{\sqrt{\mathbb{E}X^2\,\mathbb{E}Y^2}} = \frac{|XY|}{\sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}Y^2}}$$

$$\frac{a^2+b^2}{2} = \left(\frac{|X|^2}{\mathbb{E}X^2} + \frac{|Y|^2}{\mathbb{E}Y^2}\right)\frac{1}{2}$$

$$\mathbb{E}(ab) \le \mathbb{E}\left(\frac{a^2+b^2}{2}\right)$$

$$\Rightarrow \quad \frac{\mathbb{E}|XY|}{\sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}Y^2}} \le \frac{\frac{\mathbb{E}|X|^2}{\mathbb{E}X^2} + \frac{\mathbb{E}|Y|^2}{\mathbb{E}Y^2}}{2} = \frac{1+1}{2} = 1$$

$\Rightarrow$

$$\frac{\mathbb{E}|XY|}{\sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}Y^2}} \le 1$$

$$\Rightarrow \quad \mathbb{E}|XY| \le \sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}Y^2}.$$

---

② For any X, Y, set $U = X - \mathbb{E}X$

$V = Y - \mathbb{E}Y$

Then $\left|\mathbb{E}\,UV\right| \le \sqrt{\mathbb{E}U^2}\sqrt{\mathbb{E}V^2}$ by C-S

$$\Leftrightarrow \quad -1 \le \frac{\mathbb{E}\,UV}{\sqrt{\mathbb{E}U^2}\sqrt{\mathbb{E}V^2}} \le 1$$

corr(X,Y).

$$\Leftrightarrow \quad -1 \le \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}X}\sqrt{\mathrm{Var}Y}} \le 1$$

Theorem (Hölder's inequality)

For any two rvs X and Y and any $p, q \geq 1$ such that $1/p + 1/q = 1$, we have

$$|\mathbb{E}XY| \leq \mathbb{E}|XY| \leq (\mathbb{E}|X|^p)^{1/p}(\mathbb{E}|Y|^q)^{1/q}$$

The Cauchy-Schwarz is a special case of Hölder's with $p = q = 2$.

Use fact that for any $a > 0$, $b > 0$ and $p, q \geq 1$ such that $1/p + 1/q = 1$, we have

$$\frac{a^p}{p} + \frac{b^q}{q} \geq ab.$$

$$Z_1 - Z_2 = \underbrace{Z_1}_{X} + \underbrace{(-Z_2)}_{Y}$$

Theorem (Minkowski's inequality)

For any rvs X and Y and any $p \in [1, \infty)$ we have

$$(\mathbb{E}|X + Y|^p)^{1/p} \leq (\mathbb{E}|X|^p)^{1/p} + (\mathbb{E}|Y|^p)^{1/p}.$$

$p = 2$

$\sqrt{\mathrm{Var}(Z_1 - Z_2)}$

Exercise: Let Z_1 and Z_2 be standard Normals rvs.

1. Give an upper bound for $(\mathbb{E}|Z_1 - Z_2|^2)^{1/2}$ using Minkowski's inequality.
2. Give $(\mathbb{E}|Z_1 - Z_2|^2)^{1/2}$ exactly, letting $\rho = \mathrm{corr}(Z_1, Z_2)$.

① $(\mathbb{E}|Z_1 - Z_2|^2)^{1/2} \leq (\mathbb{E}|Z_1|^2)^{1/2} + (\mathbb{E}|-Z_2|^2)^{1/2} = 1 + 1 = 2$

② $(\mathbb{E}|Z_1 - Z_2|^2)^{1/2} = \sqrt{\mathrm{Var}(Z_1 - Z_2)} = \sqrt{\mathrm{Var}\, Z_1 + \mathrm{Var}\, Z_2 - 2\,\mathrm{Cov}(Z_1, Z_2)}$

$$= \sqrt{1 + 1 - 2\rho}$$

$$= \sqrt{2 - 2\rho}$$

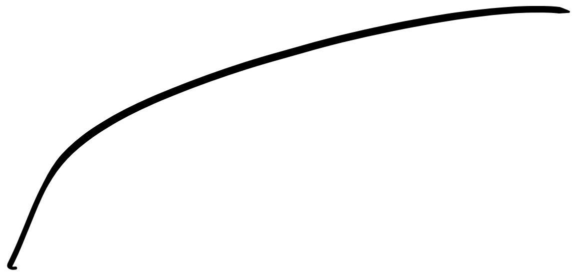
Theorem (Jensen's inequality)

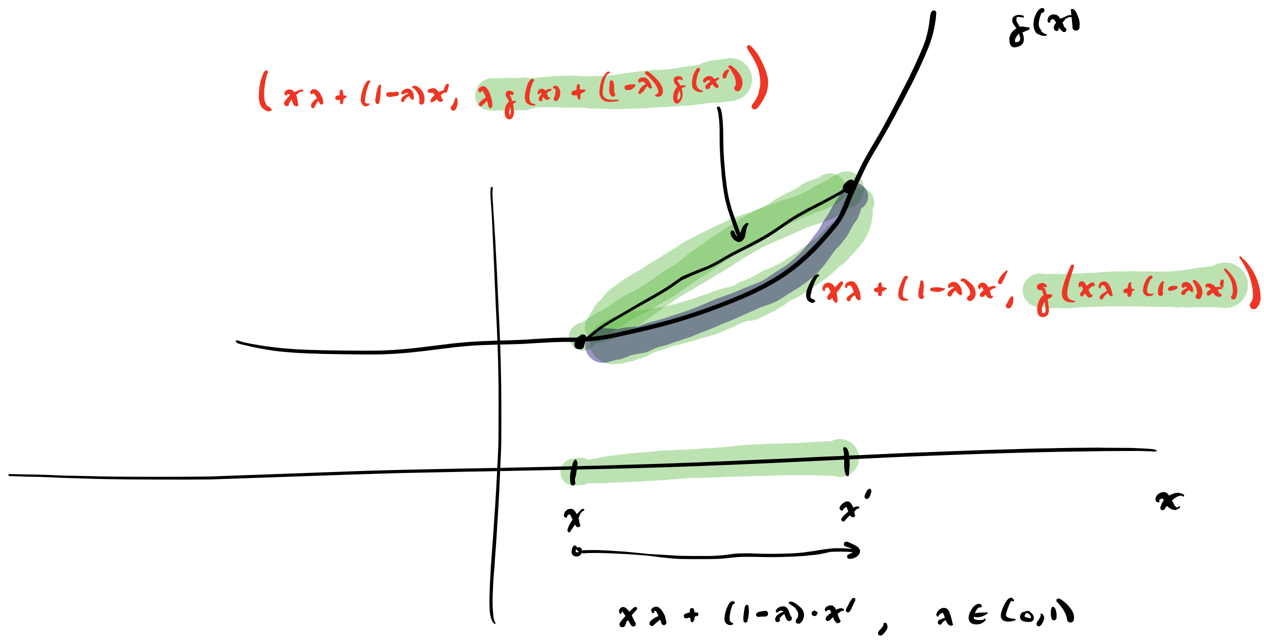
For any rv X and any convex function g we have $g(\mathbb{E}X) \leq \mathbb{E}g(X)$.

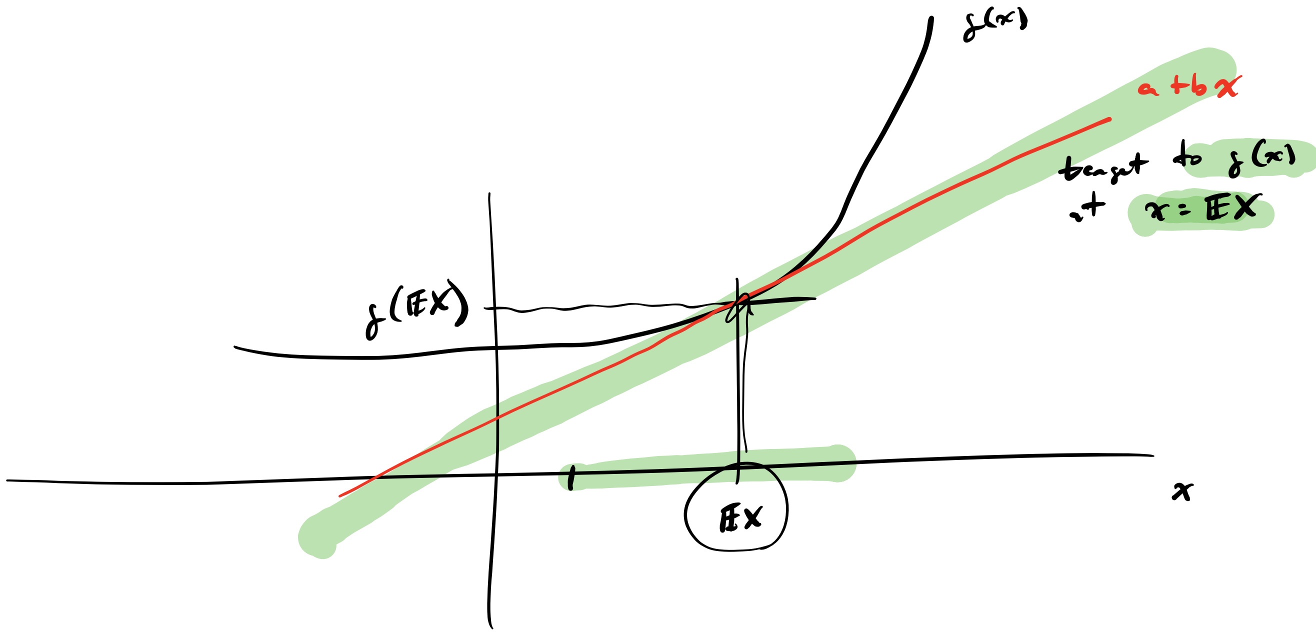
Exercise: Prove Jensen's inequality. Recall that g is convex if

$$g(\lambda x + (1-\lambda)x') \leq \lambda g(x) + (1-\lambda)g'(x) \quad \text{for all } \lambda \in (0,1),\ x, x' \in \mathbb{R}.$$

We often use Jensen's to claim $e^{\mathbb{E}X} \leq \mathbb{E}e^{X}$ and $\log \mathbb{E}X \geq \mathbb{E}\log X$.







Always have $g(X) \geq a + b X$

$$\Rightarrow \quad \mathbb{E}\, g(X) \geq a + b\, \mathbb{E}X = g(\mathbb{E}X)$$

$$P(A \cap B) = P(A|B)\,P(B)$$

1. Covariance and correlation

2. Cauchy-Schwarz, Hölder's, Minkowski's, and Jensen's inequalities

3. Hierarchical models

$$f(x,y) = f(x|y)\,f_Y(y), \qquad p(x,y) = p(x|y)\,p_Y(y)$$

- Often (X, Y) relation is most clearly described by a conditional and marginal:

$$f(y|x) = \frac{f(x,y)}{f_X(x)} \iff f(x,y) = f(y|x)f_X(x).$$

- A *hierarchical model* describes the joint dist. of an rv pair (X, Y) in the form

$$\begin{aligned} Y|X &\sim \text{Some distribution depending on } X \\ X &\sim \text{Some distribution} \end{aligned}$$

- Can use to get interesting marginal distributions for Y, which take the form

$$f_Y(y) = \int_{-\infty}^{\infty} \underbrace{f(y|x)f_X(x)}_{f(x,y)}dx.$$

Poisson-Binomial hierarchical model example

Let

$X = \#$ customers entering a store in a day

$Y = \#$ customers who make purchases

We might assume the following hierarchical model for Y:

$$Y|X \sim \text{Binomial}(X, p)$$
$$X \sim \text{Poisson}(\lambda).$$

$p_{Y|X}(y|x) = \binom{x}{y} p^y (1-p)^{x-y}$

$p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$

Exercise: Find the following:

1. The joint pmf of (X, Y).
2. The marginal pmf of Y.
3. $\boxed{\mathbb{E}Y \text{ and } \operatorname{Var} Y.}$

$\mathbb{E}Y = \mathbb{E}(\mathbb{E}[Y|X]) = \mathbb{E}(Xp) = p\mathbb{E}X = p\lambda$

$\operatorname{Var} Y = \mathbb{E}(\operatorname{Var}[Y|X]) + \operatorname{Var}(\mathbb{E}[Y|X])$
$= \mathbb{E}(Xp(1-p)) + \operatorname{Var}(Xp) = \lambda p(1-p) + p^2\lambda$
$= p\lambda.$

(1) $$p(x,y) = P(X=x \cap Y=y)$$

$$= P(Y=y \mid X=x)\, P(X=x)$$

$$= \binom{x}{y} p^y (1-p)^{x-y} \mathbb{1}(y \in \{0,1,\dots,x\}) \frac{e^{-\lambda} \lambda^x}{x!} \mathbb{1}(x \in \{0,1,2,\dots\})$$

(2) For $y = 0,1,2,\dots$

$$p_Y(y) = \sum_{x=y}^{\infty} \binom{x}{y} p^y (1-p)^{x-y} \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \sum_{x=y}^{\infty} \frac{\cancel{x!}}{(x-y)!\, y!} p^y (1-p)^{x-y} \frac{e^{-\lambda}\lambda^x}{\cancel{x!}}$$

$m = x-y,$
$x = m+y$

$$= \sum_{m=0}^{\infty} \frac{1}{m!\, y!} p^y (1-p)^m e^{-\lambda} \lambda^{m+y} \qquad e^{-\lambda} = e^{-(1-p)\lambda}\, e^{-p\lambda}$$

$$= e^{-p\lambda} \frac{(p\lambda)^y}{y!} \underbrace{\sum_{m=0}^{\infty} e^{-(1-p)\lambda} \frac{[(1-p)\lambda]^m}{m!}}_{\text{Sum over Poisson}((1-p)\lambda) = 1}$$

$$= \frac{e^{-p\lambda}(p\lambda)^y}{y!},$$

So $Y \sim \text{Poisson}(p\lambda)$

(3) $\mathbb{E}\, Y = p\lambda$

$\text{Var}\, Y = p\lambda$

W.R.T. marginal of X

Theorem (iterated expectation and iterated variance)

For any random variables X and Y we have

with respect to conditional dist. of $Y|X$

- $\mathbb{E}Y = \mathbb{E}(\mathbb{E}[Y|X])$
- $\text{Var}\, Y = \mathbb{E}(\text{Var}[Y|X]) + \text{Var}(\mathbb{E}[Y|X])$

Memorize

Exercise:

1. Prove the 1st result above.
2. Apply results to previous example.

(1) Suppose X, Y continuous

$$\mathbb{E}Y = \int_{\mathbb{R}} y\, f_Y(y)\, dy$$

$$= \int_{\mathbb{R}} y \overset{= f_Y(y)}{\left(\int_{\mathbb{R}} f(x,y)\, dx \right)} dy$$

$$= \int_{\mathbb{R}} y \int_{\mathbb{R}} f(y|x)\, f_X(x)\, dx \quad dy$$

$$= \int_{\mathbb{R}} \underbrace{\int_{\mathbb{R}} y\, f(y|x)\, dy}_{\mathbb{E}[Y|X=x]} \; f_X(x)\, dx$$

$$= \int_{\mathbb{R}} \mathbb{E}[Y|X=x]\, f_X(x)\, dx$$

$$= \mathbb{E}\left(\mathbb{E}[Y|X] \right)$$

Normal random-effects hierarchical model

Let an elementary-school pupil from the U.S. be chosen at random and let

$Y =$ score on a standardized test the selected pupil.

$A =$ average test score at the school of the selected pupil.

We might assume the following hierarchical model for Y:

$$Y|A \sim \text{Normal}(A, \sigma^2)$$
$$A \sim \text{Normal}(\mu_A, \sigma_A^2).$$

Exercise:

1. Find $\mathbb{E}Y$ and Var Y.
2. Find the marginal pdf of Y.
3. Find pdf of $A|Y$. You get $A|Y \sim \text{Normal}\left(\frac{\sigma_A^2 Y + \sigma^2 \mu_A}{\sigma_A^2 + \sigma^2}, \frac{\sigma^2 \sigma_A^2}{\sigma_A^2 + \sigma^2}\right)$

1

$$\mathbb{E}\, Y = \mathbb{E}\left(\mathbb{E}[Y|A]\right) = \mathbb{E}\left(A \right) = \mu_A$$

$$\mathrm{Var}\, Y = \mathbb{E}\left(\mathrm{Var}[Y|A]\right) + \mathrm{Var}\left(\underline{\mathbb{E}[Y|A]}\right)$$

$$= \mathbb{E}\left(\underline{\sigma^2}\right) + \mathrm{Var}\left(A \right)$$

$$= \sigma^2 + \sigma_A^2$$

2

$$f_Y(y) = \int_{\mathbb{R}} f(y|a)\, f_A(a)\, da$$

$$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{(y-a)^2}{2\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_A} e^{-\frac{(a-\mu_A)^2}{2\sigma_A^2}}\, da$$

$\vdots$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2+\sigma_A^2}} e^{-\frac{(y-\mu_A)^2}{2(\sigma^2+\sigma_A^2)}}$$

We can use hierarchical models to make interesting "mixture" distributions:

Mixture distribution induced by a hierarchical model

For a pair of rvs (X, Y), let p and f represent pmfs and pdfs, respectively. Then

- $Y|X \sim f(y|x)$ with $X \sim f_X(x)$ gives $f_Y(y) = \int_{\mathbb{R}} f(y|x) f_X(x) dx$.
- $Y|X \sim p(y|x)$ with $X \sim p_X(x)$ gives $p_Y(y) = \sum_{x \in \mathcal{X}} p(y|x) p_X(x) dx$.
- $Y|X \sim f(y|x)$ with $X \sim p_X(x)$ gives $f_Y(y) = \sum_{x \in \mathcal{X}} f(y|x) \boxed{p_X(x)} dx$
- $Y|X \sim p(y|x)$ with $X \sim f_X(x)$ gives $p_Y(y) = \int_{\mathbb{R}} p(y|x) f_X(x) dx$.

The marginal distributions of Y are called *mixture distributions*.

Beta-Binomial hierarchical model example

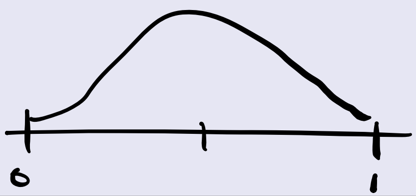
Let a spectator of a basketball game be chosen at random and let

$Y = \#$ freethrows made out of n attempts by the chosen spectator.

$P =$ free-throw success rate of the chosen spectator.

We might assume the following hierarchical model for Y:

$$Y|P \sim \text{Binomial}(n, P)$$
$$P \sim \text{Beta}(\alpha, \beta).$$



Exercise: Find

1. $\mathbb{E}Y$ and $\text{Var}\, Y$.
2. The marginal distribution of Y.

① $\mathbb{E}\, Y = \mathbb{E}\left(\mathbb{E}[Y|P] \right) = \mathbb{E}\left(nP \right) = n\,\mathbb{E}\,P = n \frac{\alpha}{\alpha+\beta}$

$$\begin{aligned}
\mathrm{Var}\, Y &= \mathbb{E}\left(\mathrm{Var}[Y|P] \right) + \mathrm{Var}\left(\mathbb{E}[Y|P] \right) \\
&= \mathbb{E}\left(n P(1-P) \right) + \mathrm{Var}\left(n\,P \right) \\
&= n \left(\mathbb{E}\, P - \mathbb{E}\, P^2 \right) + n^2\, \mathrm{Var}\, P
\end{aligned}$$

$n \quad \frac{\alpha}{\alpha+\beta} \qquad \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)} + \left(\frac{\alpha}{\alpha+\beta}\right)^2 \qquad \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}$

$$= n \left[\frac{\alpha}{\alpha+\beta} - \left(\frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)} + \left(\frac{\alpha}{\alpha+\beta}\right)^2 \right) \right] + n^2 \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}$$

②

$$p_Y(y) = \int_{\mathbb{R}} p(y|p)\, f_P(p)\, dp$$

$$= \int_0^1 \binom{n}{y} p^y (1-p)^{n-y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}\, dp$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \binom{n}{y} \frac{\Gamma(y+\alpha)\Gamma(n+\beta-y)}{\Gamma(\alpha+n+\beta)} \underbrace{\int_0^1 \frac{\Gamma(\alpha+n+\beta)}{\Gamma(y+\alpha)\,\Gamma(n+\beta-y)} p^{y+\alpha-1} (1-p)^{n+\beta-y-1}\, dp}_{=1}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \binom{n}{y} \frac{\Gamma(y+\alpha)\Gamma(n+\beta-y)}{\Gamma(\alpha+n+\beta)} \qquad \text{for} \quad y = 0, 1, \dots, n.$$

"Beta-Binomial" dist.

$$\mathbb{E}[XY \mid X=x] = \int xy\, f(y|x)\, dy = x \int y\, f(y|x)\, dy$$
$$= x\, \mathbb{E}[Y \mid X=x]$$

Exercise: Let (X, Y) be a pair of rvs such that

$$\frac{\alpha}{\alpha+\beta}$$

$$Y|X \sim \text{Beta}(3/2 - X, 1/2 + X)$$
$$\boxed{X \sim \text{Uniform}(0,1)}$$

1. Find $\text{Cov}(X, Y) = \underline{\mathbb{E}XY - \underbrace{\mathbb{E}X}_{\frac{1}{2}}\, \mathbb{E}Y}$
2. Find $\text{Var}\, Y$.

$$\mathbb{E}Y = \mathbb{E}\left(\mathbb{E}[Y|X]\right)$$
$$= \mathbb{E}\left(\frac{3/2 - X}{3/2 - X + 1/2 + X}\right)$$
$$= \mathbb{E}\left(\frac{3/2 - X}{2}\right)$$
$$= \frac{3/2 - 1/2}{2}$$
$$= \frac{1}{2}$$

We have

$$\underline{\mathbb{E}XY} = \mathbb{E}\left(\mathbb{E}[XY|X]\right)$$
$$= \mathbb{E}\left(X\, \mathbb{E}[Y|X]\right)$$
$$= \mathbb{E}\left(X \cdot \frac{3/2 - X}{3/2 - X + 1/2 + X}\right)$$

$$= \mathbb{E}\left(x \cdot \frac{\left(\frac{3}{2} - x\right)}{2} \right)$$

$$= \mathbb{E}\left[x \cdot \frac{3}{4} - \frac{x^2}{2} \right]$$

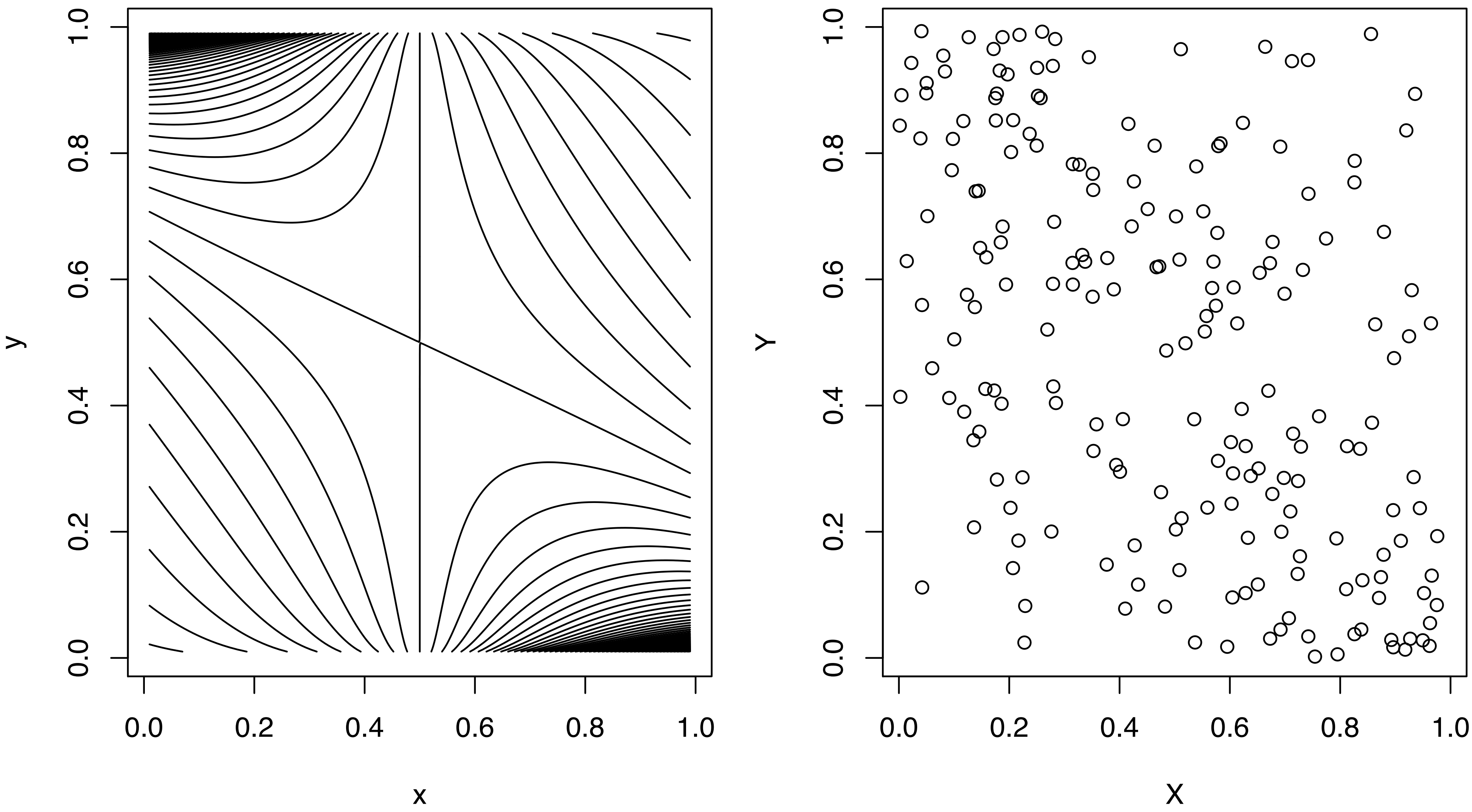
$$= \frac{1}{2} \frac{3}{4} - \frac{1}{3} \cdot \frac{1}{2}$$

$$= \frac{3}{8} - \frac{1}{6}$$

$$= \frac{9 - 4}{24}$$

$$= \frac{5}{24}$$

$$\mathrm{Cov}(X, Y) = \frac{5}{24} - \frac{1}{2} \cdot \frac{1}{2} = \frac{5}{24} - \frac{1}{4} = \frac{5}{24} - \frac{6}{24} = -\frac{1}{24}.$$


0.0
0.2
0.4
0.6
0.8
1.0
x
y
X
Y

Multinoulli trial

A *multinoulli trial* is an experiment in which there are K possible outcomes which occur with the probabilities $p_1, \ldots, p_K$, where $\sum_{k=1}^K p_k = 1$.

Extends Bernoulli trial (two outcomes) to two or more outcomes.

Multinoulli distribution

Let the random variables $X_1, \dots, X_K$ encode the outcome of a multinoulli trial as

$$X_k = \begin{cases} 1 & \text{if outcome } k \text{ occurs} \\ 0 & \text{otherwise} \end{cases} \quad \text{for } k = 1, \dots, K.$$

Then the set $(X_1, \dots, X_K)$ of K rvs has the *multinoulli distribution* and we write

$$(X_1, \dots, X_K) \sim \text{Multinoulli}(p_1, \dots, p_K).$$

Exercise:

1. Write down the joint pmf of $(X_1, \dots, X_K)$.
2. Find the marginal pmf of X_k for each $k = 1, \dots, K$.
3. Find $\mathbb{E}X_k$ and $\text{Cov}(X_k, X_{k'})$.

$$(X_1, \dots X_K) \in \begin{cases} (1,0,0\dots,0) & \leftarrow p_1 \\ (0,1,0\dots 0) & \leftarrow p_2 \\ (0,0,1,\dots 0) & \leftarrow p_3 \\ \vdots & \\ (0,0,0\dots 1) & \leftarrow p_K \end{cases}$$

$$p(x_1, \dots, x_K) = p_1^{x_1} \cdot p_2^{x_2} \cdot \dots \cdot p_K^{x_K} \cdot \mathbb{1}\left(x_1, \dots x_K \in \{0,1\},\ \sum_{k=1}^{K} x_k = 1 \right)$$

For any $k = 1, \dots, K$

$X_k \in \{0,1\}$

② $$p_{X_k}(x_k) = \begin{cases} \displaystyle\sum_{\substack{x_1,\dots,x_K \in \{0,1\},\ \sum_{k=1} x_k = 1 \\ x_k = 1}} p_1^{x_1} \cdot p_2^{x_2} \cdot \dots \cdot p_K^{x_K} = p_k & x_k = 1 \\ 1 - p_k & x_k = 0 \end{cases}$$

③ $\mathbb{E} X_k = p_k$

$k \neq j$ $$\mathrm{Cov}(X_k, X_j) = \mathbb{E} X_k X_j - \mathbb{E} X_k \mathbb{E} X_j$$

$$= 0 - p_k p_j$$

$$= -p_k p_j$$

Gaussian mixture hierarchical model

Consider the hierarchical model

$$Y|(X_1,\dots,X_K) \sim \text{Normal}\left(\sum_{k=1}^K X_k\mu_k, \sum_{k=1}^K X_k\sigma_k^2\right) \quad \text{and}$$
$$(X_1,\dots,X_K) \sim \text{Multinoulli}(p_1,\dots,p_K),$$

where $(\mu_1,\sigma_1^2),\dots,(\mu_K,\sigma_K^2)$ are K mean and variance pairs.

Exercise: Find $\mathbb{E}Y$ and $\text{Var}\, Y$. Then find the marginal distribution of Y.

$$\mathbb{E}Y = \mathbb{E}\left(\mathbb{E}\left[Y|(X_1,\dots,X_K)\right]\right) = \mathbb{E}\left(\sum_{k=1}^K X_k\mu_k\right) = \sum_{k=1}^K p_k\mu_k$$

$$\text{Var}\, Y = \mathbb{E}\left(\text{Var}\left[Y|X_1\dots X_K\right]\right) + \text{Var}\left(\mathbb{E}\left[Y|X_1\dots X_K\right]\right)$$

$$= \mathbb{E}\left(\sum_{k=1}^{K} X_k \sigma_k^2\right) + \mathrm{Var}\left(\sum_{k=1}^{K} X_k \mu_k\right)$$

$$= \underbrace{\sum_{k=1}^{K} p_k \sigma_k^2}_{\text{weighted avg of variances}} + \sum_{k=1}^{k} \mu_k^2 \underbrace{\mathrm{Var}\, X_k}_{p_k(1-p_k)} + \sum_{k \neq j} \mu_k \mu_j \underbrace{\mathrm{Cov}(X_k, X_j)}_{-p_k p_j}$$

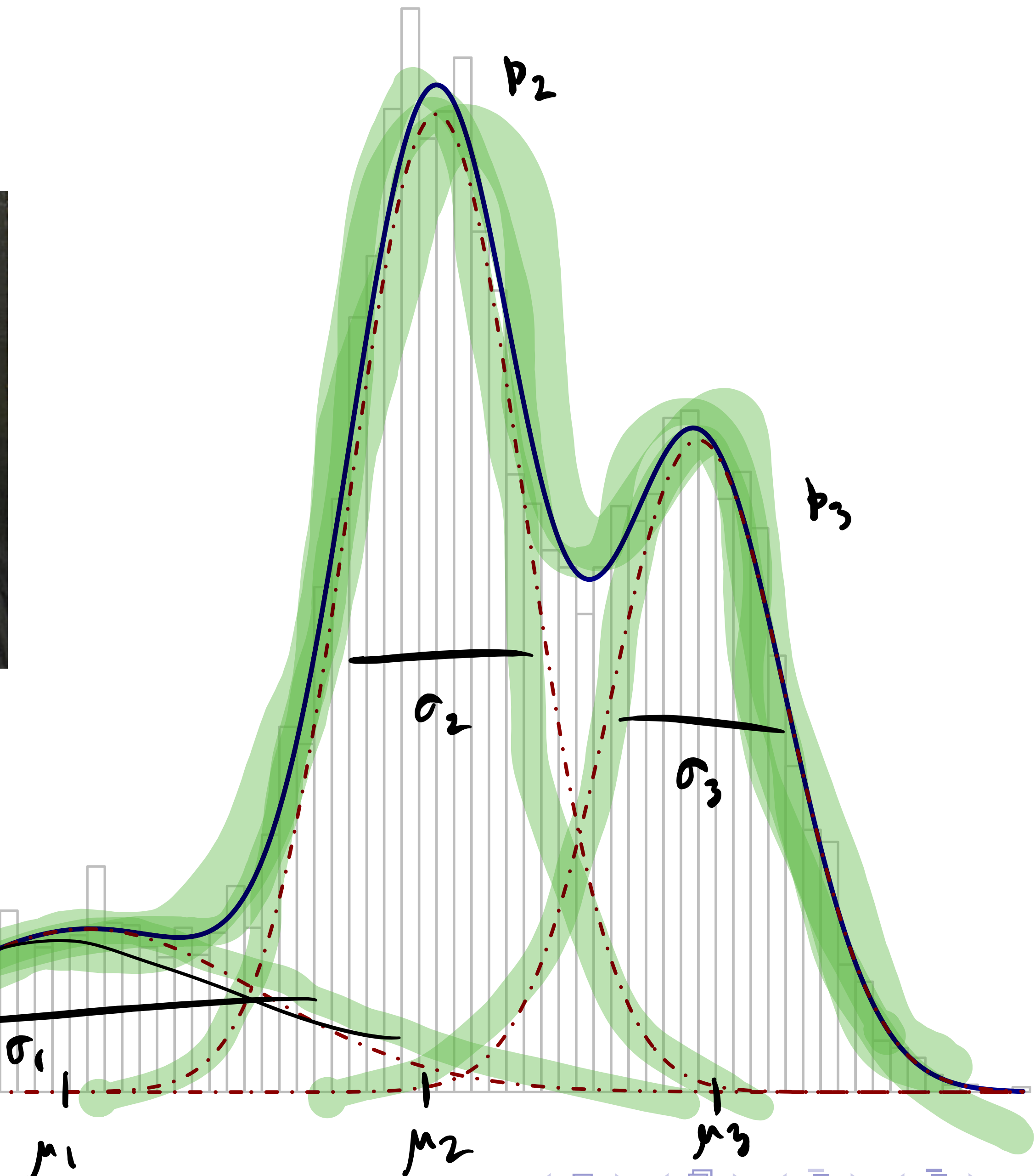
$$= \underbrace{\sum_{k=1}^{K} p_k \sigma_k^2}_{\text{weighted avg of variances}} + \sum_{k=1}^{k} \mu_k^2 p_k (1-p_k) - \sum_{k \neq j} \mu_k \mu_j p_k p_j$$

$$= \underbrace{\sum_{k=1}^{K} p_k \sigma_k^2}_{\text{weighted mean of variances}} + \underbrace{\sum_{k=1}^{k} p_k \left(\mu_k - \sum_{j=1}^{K} p_j \mu_j\right)^2}_{\text{weighted variance of means}}$$

$$\mathrm{Var}\left(\sum a_i X_i\right) = \sum a_i^2 \mathrm{Var}\, X_i + \sum_{i \neq j} a_i a_j \mathrm{Cov}(X_i, X_j)$$

Marginal pdf of Y:

$$f_Y(y) = \sum_{k=1}^{K} p_k \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_k} e^{-\frac{(y-\mu_k)^2}{2\sigma_k^2}}$$


p_1
p_2
p_3
σ_1
σ_2
σ_3
μ_1
μ_2
μ_3

Multinomial distribution

For $k = 1, \dots, K$, let Y_k be # times outcome k occurrs in n independent Multinoulli trials with the outcome probabilities $p_1, \dots, p_K$.

Then the set of rvs $(Y_1, \dots, Y_K)$ has the *multinomial distribution* and we write

$$(Y_1, \dots, Y_K) \sim \text{Multinomial}(n, p_1, \dots, p_K).$$

The joint pmf of $(Y_1, \dots, Y_K)$ is given by

$$p(y_1, \dots, y_K) = \left(\frac{n!}{y_1! \cdots y_K!} \right) p_1^{y_1} \cdots p_K^{y_K}$$

for $(y_1, \dots, y_K) \in \{0, 1, \dots, n\}^K$ st $\sum_{k=1}^K y_k = n$.

Note: For $k = 1, \dots, K$ the marginal distribution of y_k is $\text{Binomial}(n, p_k)$.

Exercise: Let $Z \sim \text{Normal}(0,1)$ and let $\mu \in \mathbb{R}$.

1. Find the pdf of $W = (Z+\mu)^2$. This has the *non-central chi-squared dist*.
2. Find the marginal pdf of U in the hierarchical model

$$\begin{aligned} U|K &\sim \chi^2_{1+2K} \\ K &\sim \text{Poisson}(\mu^2/2). \end{aligned}$$

3. Compare these using the fact that

$$\frac{2^k}{\sqrt{2\pi}}\frac{k!}{(2k)!} = \frac{1}{\Gamma(k+1/2)2^{k+1/2}} \quad \text{for all } k = 0,1,\ldots$$