Mathematics/Statistics Bootcamp

Part III: Distributions

Steven Winter Christine Shen

Department of Statistical Science
Duke University

MSS Orientation, August 2022

Outline

Random Variables

Random Variables

A **random variable** is a (measurable) function from a sample space to an outcome space ($\mathbb{R}$, $\mathbb{Z}$, sentences, brain scans, etc).

Common notation:

- Ω is the set of all possible outcomes of an experiment.
- $\omega \in \Omega$ is a particular outcome.
- $Y = Y(\omega)$ is a function of ω (the random variable).

Imagine our experiment is rolling two dice. What is Ω? Give an example of ω and a few random variables.

Cumulative Distribution Functions

The **cumulative distribution function** (CDF) is

$$F_X(x) = P_X(X \leq x) \qquad \forall x \in \mathbb{R}$$

A function $F : \mathbb{R} \to [0, 1]$ is a CDF if and only if the following are true:

- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.
- $F(x)$ is non-decreasing.
- $F(x)$ is right continuous.

A random variable is **continuous** if $F_X(x)$ is continuous. A random variable is **discrete** if $F_X(x)$ is a step function.

Example

Consider flipping a biased coin. Let p denote the probability of getting a head. If we define

$$X = \text{number of tosses until a head},$$

then for $x \in \mathbb{N}$,

$$\begin{aligned} P(X \leq x) &= p + (1-p)p + (1-p)^2 p + \cdots + \cdots (1-p)^{x-1} p \\ &= \sum_{i=1}^{x} (1-p)^{i-1} p \\ &= 1 - (1-p)^x. \end{aligned}$$

Probability Density/Mass Functions

The **probability mass function** (PMF) for a discrete random variable is

$$f_X(x) = P(X = x); \quad \forall x \in \Omega.$$

The **probability density function** (pdf) for an (absolutely) continuous random variable is a function f_X such that

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t; \quad \forall x \in \Omega.$$

Recall PMFs/PDFs sum/integrate to 1. Can help you avoid calculus.

Univariate Distributions

Bernoulli Distribution

Represents a single coin flip with success $(X = 1)$ probability p.

$$
\begin{aligned}
X &\sim \mathrm{Bernoulli}(p) \\
P(X = x) &= p^x(1-p)^{1-x}, \quad x \in \{0, 1\} \\
\mathbb{E}[X] &= p \\
\mathbb{V}[X] &= p(1-p)
\end{aligned}
$$

Binomial Distribution

Counts the number of successes in n independent trials all with the same success probability p.

$$
\begin{aligned}
X &\sim \mathrm{Binomial}(n, p) \\
P(X = x) &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \quad x \in \{0, 1, \ldots, n\} \\
\mathbb{E}[X] &= np \\
\mathbb{V}[X] &= np(1-p)
\end{aligned}
$$

Can decompose $X = \sum_{i=1}^{n} Y_i$ where the Y_i are iid $\mathrm{Bernoulli}(p)$.

Poisson Distribution

Counts events occurring with rate λ:

$$X \sim \mathrm{Poisson}(\lambda)$$
$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x \in \{0, 1, 2, \ldots\}$$
$$\mathbb{E}[X] = \mathbb{V}[X] = \lambda$$

Not always a great fit: real data are often zero-inflated and over-dispersed.

If $X \sim Po(\lambda)$ and $Y \sim Po(\eta)$ are independent, then $X + Y \sim Po(\lambda + \eta)$.

Geometric Distribution

Counts the number of failures until the first success in sequential independent Bernoulli trials with success probability p.

$$\begin{aligned} X &\sim \mathrm{Geom}(p) \\ P(X = k) &= (1-p)^k p, \quad k \in \{0, 1, 2, \ldots\} \\ \mathbb{E}[X] &= \frac{1-p}{p} \\ \mathbb{V}[X] &= \frac{1-p}{p^2} \end{aligned}$$

Caution: some parameterizations start at 1.

Normal Distributions

A random variable $X \sim N(\mu, \sigma^2)$ has PDF:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(x-\mu)^2}{2\sigma^2} \right)$$

Sometimes write in terms of **precision** (inverse variance): $X \sim N(\mu, \phi^{-1})$.

If $Z \sim N(0, 1)$ then the distribution of Z is **standard normal.**

Normal Distributions Properties

Transformations: If $X \sim N(\mu, \sigma^2)$, then

$$a + bX \sim N(a + b\mu, b^2\sigma^2).$$

Decompositions: Can always decompose $X \sim N(\mu, \sigma^2)$ as $X = \mu + \sigma Z$ with $Z \sim N(0, 1)$.

Sums: If $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\eta, \tau^2)$ are jointly normal, then

$$X + Y \sim N(\mu + \eta, \sigma^2 + \tau^2 + 2\mathrm{Cov}[X, Y]).$$

Independence: *Jointly* normal random variables are independent if and only if they are uncorrelated.

Chi-Squared Distribution

If Z_1, $Z_2, \ldots, Z_k$ are independent, standard normal random variables, then

$$\sum_{j=1}^{k} Z_j^2 \sim \chi_k^2$$

follows a Chi-Squared distribution with k degrees of freedom. Special case of the Gamma distribution.

You will see statements like $X \sim c\chi_k^2$ in linear models. This means $X/c \sim \chi_k^2$.

Gamma Distribution

A random variable $X \sim \mathrm{Gamma}(\alpha, \beta)$ has PDF:

$$
\begin{aligned}
f_X(x) &= \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-x\beta\} \\
\mathbb{E}[X] &= \frac{\alpha}{\beta} \\
\mathbb{V}[X] &= \frac{\alpha}{\beta^2} \\
&\alpha, \beta > 0 \\
&x \in (0, \infty)
\end{aligned}
$$

This is the rate parameterization. You will also see the scale parameterization with $\theta = 1/\beta$.

If $X \sim \mathrm{Gamma}(1, \beta)$, then $X \sim \mathrm{Exponential}(\lambda = \beta)$.

Gamma Distribution Properties (Optional)

Useful facts:

- If $X \sim \mathrm{Gamma}(\alpha_1, \beta)$ and $Y \sim \mathrm{Gamma}(\alpha_2, \beta)$ are independent, then $X + Y \sim \mathrm{Gamma}(\alpha_1 + \alpha_2, \beta)$. Consequence: sum of independent exponentials is Gamma.
- If $\alpha = \frac{\nu}{2}$ and $\beta = \frac{1}{2}$ then $X \sim \chi^2_\nu$.
- If $X \sim \mathrm{Gamma}(\alpha_1, \beta)$ and $Y \sim \mathrm{Gamma}(\alpha_2, \beta)$ are independent, then $X/(X + Y) \sim \mathrm{Beta}(\alpha_1, \alpha_2)$.
- Let $X_i \sim \mathrm{Gamma}(\alpha_i, \beta)$, $i = 1, ..., n$ be independent. Set $T = \sum_i X_i$. Then $(X_1/T, ..., X_n/T) \sim \mathrm{Dirichlet}(\alpha_1, ...\alpha_n)$.
- If $X \sim \mathrm{Gamma}(\alpha, \beta)$, then $\frac{1}{X} \sim \mathrm{Inverse-Gamma}(\alpha, \beta)$.

Student's-t Distribution

A random variable T follows a Student's-t distribution with ν degrees of freedom if

$$\begin{aligned} T &= \frac{Z}{\sqrt{V/\nu}}, \\ Z &\sim N(0,1), \\ V &\sim \chi^2_\nu \end{aligned}$$

and Z and V are independent.

Like a normal, but with heavier tails. The $\nu = 1$ case is a Cauchy distribution with undefined mean and variance.

As $\nu \to \infty$, $T \to N(0,1)$.

Useful for confidence intervals with unknown variance.

F Distribution

A random variable X follows a F-distribution with numerator degrees of freedom ν_1 and denominator degrees of freedom ν_2 if

$$X = \frac{V_1/\nu_1}{V_2/\nu_2}$$

where V_1 and V_2 are independent χ^2 random variables with degrees of freedom equal to ν_1 and ν_2 respectively.

Useful for model selection, e.g. comparing variances in ANOVA.

Beta Distribution

A random variable $X \sim \mathrm{Beta}(\alpha, \beta)$ has PDF:

$$
\begin{aligned}
f_X(x) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in (0,1), \quad \alpha, \beta > 0 \\
\mathbb{E}[X] &= \frac{\alpha}{\alpha+\beta} \\
\mathbb{V}[X] &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}
\end{aligned}
$$

Useful for eliciting probability distributions for proportions.

Exercises

1. Give an example of a random variable which is not discrete or continuous.
2. Let $X_1, ..., X_n$ be iid $N(\mu, \sigma^2)$. Find the distributions of
   a) $\bar{X} = \sum_{i=1}^{n} X_i/n$.
   b) $(X_i - \bar{X})^2$. Recall $Cov[\cdot, \cdot]$ is linear in both components.
   c) $(X_i - \mu)/(X_j - \mu)$
3. A $\mathrm{Weibull}(\alpha, \beta)$ distribution has density

$$f(x) = \alpha\beta(\beta x)^{\alpha-1} \exp(-(\beta x)^{\alpha}).$$

where $\alpha, \beta, x > 0$. Find the CDF.

Solutions

1. Many answers - e.g., let $X \sim Po(\lambda)$ and $Y \sim N(\mu, \sigma^2)$ be independent, then set $Z = X + Y$.
2. a) Since the X_i are independent, $\bar{X} \sim N(\mu, \sigma^2/n)$.

   b) First calculate $Cov[X_i, \bar{X}] = \sigma^2/n$. Then

   $$X_i - \bar{X} \sim N\left(0, \sigma^2 + \frac{\sigma^2}{n} - \frac{2\sigma^2}{n}\right) = N\left(0, \frac{n-1}{n}\sigma^2\right)$$

   Therefore $(X_i - \bar{X})^2 \sim ((n-1)\sigma^2/n)\chi_1^2$.

   c) If $i = j$ then this is a constant variable equal to 1. If $i \neq j$,

   $$\frac{X_i - \mu}{X_j - \mu} = \frac{(X_i - \mu)/\sigma}{(X_j - \mu)/\sigma}$$

   is a ratio of two independent standard normals, Z_1 and Z_2. In distribution, we have

   $$\frac{Z_1}{Z_2} = \frac{Z_1}{\sqrt{Z_2^2/1}} = \frac{N(0,1)}{\sqrt{\chi_1^2/1}} \sim t_1.$$

Solutions (Optional)

2. c) Problem: $Z_1/\sqrt{Z_2^2} = Z_1/|Z_2|$, **not** Z_1/Z_2! Need to check equality in distribution, i.e. $P[Z_1/Z_2 \leq x] = P[Z_1/|Z_2| \leq x]$.

   Calculate

   $$\begin{aligned} P[Z_1/Z_2 \leq x] &= P[Z_1/Z_2 \leq x, Z_2 > 0] + P[Z_1/Z_2 \leq x, Z_2 < 0] \\ &= P[Z_1/|Z_2| \leq x, Z_2 > 0] + P[-Z_1/|Z_2| \leq x, Z_2 < 0] \end{aligned}$$

   Note $-Z_1 \sim N(0,1)$ is still independent of Z_2, so $-Z_1/|Z_2|$ has the same distribution as Z_1/Z_2. Thus

   $$\begin{aligned} P[Z_1/Z_2 \leq x] &= P[Z_1/|Z_2| \leq x, Z_2 > 0] + P[Z_1/|Z_2| \leq x, Z_2 < 0] \\ &= P[Z_1/|Z_2| \leq x] \end{aligned}$$

Solutions

3. Note

$$
\begin{aligned}
\frac{d}{dx}\exp(-(\beta x)^\alpha) &= \exp(-(\beta x)^\alpha)\frac{d}{dx}\Big(-(\beta x)^\alpha\Big) \\
&= \exp(-(\beta x)^\alpha)(-\alpha(\beta x)^{\alpha-1})\frac{d}{dx}\Big(\beta x\Big) \\
&= -\alpha\beta(\beta x)^{\alpha-1}\exp(-(\beta x)^\alpha)
\end{aligned}
$$

Therefore the CDF is

$$
F(x) = \int_{-\infty}^{x} f(t)dt = 1 - \exp(-(\beta x)^\alpha)
$$

for $x > 0$ and zero otherwise.

Multivariate Distributions

Random Vectors

A d-dimensional **random vector** is a collection of d random variables:

$$\mathbf{X} = (X_1, \cdots, X_d)^T$$

The joint distribution function of the random vector $\mathbf{X}$ is

$$\begin{aligned} F_X(\mathbf{x}) &= F_X(x_1, \ldots, x_d) \\ &= P(X_1 \le x_1, \ldots, X_d \le x_d) \\ &= P(\mathbf{X} \le \mathbf{x}) \end{aligned}$$

If F_X is absolutely continuous, then the joint density function f_X of $\mathbf{X}$ is

$$f_X(\mathbf{x}) = f_X(x_1, \ldots, x_d) = \frac{\partial^d F_X(x_1, \ldots, x_d)}{\partial x_1 \cdots \partial x_d}$$

Independence

We can find the **marginal density** of a random variable by integrating/summing out the others. For example, if we have a joint bivariate density $f_{X,Y}(x,y)$, then

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

The components of a random vector $\mathbf{X}$ are **independent** if the joint CDF (equivalently PDF) factors as a product of marginals:

$$F_X(\mathbf{x}) = \prod_{i=1}^{d} F_i(x_i), \quad f_X(\mathbf{x}) = \prod_{i=1}^{d} f_i(x_i)$$

Multivariate Moments

The expected value of a random vector $\mathbf{X}$ is

$$\mu_X = E(\mathbf{X}) = (E(X_1), \cdots, E(X_d)) = (\mu_1, \cdots, \mu_d)^T$$

and the $d \times d$ **covariance matrix** is

$$\mathbf{\Sigma}_X = \mathrm{Cov}(\mathbf{X}, \mathbf{X}) = E[(\mathbf{X} - \mu_X)(\mathbf{X} - \mu_X)^T]$$

If $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$, then

$$\begin{aligned} \mu_Y &= \mathbf{A}\mu_x + \mathbf{b}, \\ \mathbf{\Sigma}_Y &= \mathbf{A}\mathbf{\Sigma}_{\mathbf{X}}\mathbf{A}^T. \end{aligned}$$

Multivariate Normal Distribution

Let $\mathbf{Z}$ be a standard normal vector - i.e., $\mathbf{Z} = (Z_1, ..., Z_n)$ where the Z_i are iid standard normal.

A random vector $\mathbf{X}$ has a multivariate normal distribution with mean vector μ and positive definite symmetric covariance matrix $\mathbf{\Sigma}$ if and only if

$$\mathbf{X} = \mu + L\mathbf{Z}$$

where $\mathbf{LL}^T = \mathbf{\Sigma}$.

We write $\mathbf{X} \sim N_n(\mu, \mathbf{\Sigma})$. The density is

$$f(\mathbf{x}) = (2\pi)^{-n/2}|\mathbf{\Sigma}|^{-1/2}e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)}.$$

Multivariate Normal Properties

If $\mathbf{X} \sim N(\mu_{\mathbf{X}}, \mathbf{\Sigma}_{\mathbf{X}})$, then

$$\mathbf{AX} + \mathbf{b} \sim N(\mathbf{A}\mu_{\mathbf{X}} + \mathbf{b}, \mathbf{A\Sigma_X A}^T)$$

Note: can always rotate $\mathbf{X}$ to make coordinates independent.

If $\mathbf{Y} \sim N(\mu_{\mathbf{Y}}, \mathbf{\Sigma}_{\mathbf{Y}})$ is independent of $\mathbf{X}$, then

$$\mathbf{X} + \mathbf{Y} \sim N(\mu_{\mathbf{X}} + \mu_{\mathbf{Y}}, \mathbf{\Sigma}_{\mathbf{X}} + \mathbf{\Sigma}_{\mathbf{Y}})$$

Sum of correlated MVN is still MVN, provided $\mathbf{X}$ and $\mathbf{Y}$ are jointly normal.

If $\mathbf{X}$ and $\mathbf{Y}$ are jointly normal, then $\mathbf{X}|\mathbf{Y}$ and $\mathbf{Y}|\mathbf{X}$ are also normal.[1]

[1] Hard exercise: derive the conditional distributions.

High Dimensional Vectors (Optional)

Let $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I}_n)$ by an n-dimensional normal vector. Samples of each component are clustered around $\mathbf{0}$.

Intuitively, we expect samples of $\mathbf{Z}$ to be clustered around $\mathbf{0}_n$. This is wrong!

In high dimensions, $\mathbf{Z}$ *concentrates around an* $n-1$ *dimensional sphere* of radius $\sqrt{n}$:

$$N(\mathbf{0}, \mathbf{I}_n) \approx Unif(\sqrt{n}\mathbb{S}^{n-1})$$

Figure from "High-Dimensional Probability" by Roman Vershynin.

Exercises

Consider the linear model $\mathbf{Y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I})$ where σ^2 is known, $\mathbf{Y} \in \mathbb{R}^n$ and $\mathbf{X} \in \mathbb{R}^{n\times p}$ (rank $p \leq n$). We found

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

minimized $||\mathbf{Y} - \mathbf{X}\beta||^2$, giving fitted values

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

Find distributions for the following:

1. $\hat{\beta}$, $\hat{\beta}_1$, and $(\hat{\beta}_1, \hat{\beta}_7)^T$ (if $p \geq 7$).
2. $\hat{\mathbf{Y}}$ and $\mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{P})\mathbf{Y}$. Guess the distribution of $\mathbf{Y}^T(\mathbf{I} - \mathbf{P})\mathbf{Y}$.
3. $(\mathbf{X}^T\mathbf{X})^{1/2}(\hat{\beta} - \beta)$ and $||\mathbf{X}(\hat{\beta} - \beta)||^2$.

What do these quantities mean?

Solutions

1. These are the estimated coefficients - i.e., our best guess at the slope β after seeing the data.

   By linearity,

   $$\begin{aligned}\hat{\beta} &\sim N((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\beta, (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1})\\ &\sim N(\beta, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}).\end{aligned}$$

   Consequently,

   $$\begin{aligned}\hat{\beta}_1 &\sim N(\beta_1, \sigma^2[(\mathbf{X}^T\mathbf{X})^{-1}]_{1,1})\\ \begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_7\end{pmatrix} &\sim N\left(\begin{pmatrix}\beta_1\\ \beta_7\end{pmatrix}, \sigma^2[(\mathbf{X}^T\mathbf{X})^{-1}]_{(1,7),(1,7)}\right)\end{aligned}$$

Solutions

2. By linearity,

$$\hat{\mathbf{Y}} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T) \sim N(\mathbf{X}\beta, \sigma^2\mathbf{P})$$

These our predicted values along the line $\mathbf{Y} = \mathbf{X}\beta$.

Can compute $\mathbf{Y} - \hat{\mathbf{Y}}$ by finding the covariance or by linearity:

$$\begin{aligned}(\mathbf{I} - \mathbf{P})\mathbf{Y} &\sim N((\mathbf{I} - \mathbf{P})\mathbf{X}\beta, \sigma^2(\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P})^T) \\ &\sim N(\mathbf{0}, \sigma^2(\mathbf{I} - \mathbf{P}))\end{aligned}$$

These are the residuals/errors of our predictions.

One can show the sum of squared residuals, $\mathbf{Y}^T(\mathbf{I} - \mathbf{P})\mathbf{Y}$, is proportional to a χ^2_{n-p}. Proof: apply spectral theorem or compact SVD to $\mathbf{I} - \mathbf{P}$.

Solutions

3. From part 1),

$$(\mathbf{X}^T\mathbf{X})^{1/2}(\hat{\beta} - \beta) \sim N(\mathbf{0}, \sigma^2\mathbf{I})$$

   Therefore $||X(\hat{\beta} - \beta)||^2 \sim \sigma^2\chi^2_p$. This is the distance our predictions are away from the true mean of the line.

Building New Variables

Univariate Change of Variables

If $Y = g(X)$ for some monotone g, then

$$F_Y(y) = P_Y[Y \leq y] = P_X[g(X) \leq y] = P_X[X \leq g^{-1}(y)]$$
$$= F_X(g^{-1}(y)).$$

Taking derivatives,

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

E.g. if $Y = g(X) = 2X$ then $f_Y(y) = f_X(y/2)/2$.

If g is not monotone, then

$$f_Y(y) = \sum_{k=1}^{n(y)} f_X(g_k^{-1}(y)) \left| \frac{d}{dy}(g_k^{-1}(y)) \right|.$$

where $g_1^{-1}(y), \ldots, g_{n(y)}^{-1}(y)$ are the $n(y)$ solutions to $g(x) = y$.

Example

Let $X \sim \mathrm{Normal}(0,1)$ and $Y = g(X) = X^2$.

If $y = g(x)$, then $x = g^{-1}(y) = \pm\sqrt{y}$. Let $g_1^{-1}(y) = -\sqrt{y}$ and $g_2^{-1}(y) = \sqrt{y}$. Almost always $n(y) = 2$. Therefore

$$
\begin{aligned}
f_Y(y) &= f_X(g_1^{-1}(y))\left|\frac{d}{dx}(g_1^{-1}(y))\right| + f_X(g_2^{-1}(y))\left|\frac{d}{dx}(g_2^{-1}(y))\right| \\
&= \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-(-\sqrt{y})^2}{2}\right)\left|-\frac{1}{2\sqrt{y}}\right| + \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-(\sqrt{y})^2}{2}\right)\left|\frac{1}{2\sqrt{y}}\right| \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{y}}\exp\left(\frac{-y}{2}\right)
\end{aligned}
$$

This is a $\mathrm{Gamma}(1/2, 1/2)$ (equivalently χ_1^2) density.

Multivariate Change of Variables

If $\mathbf{Y} = g(\mathbf{X})$ with differentiable inverse. Then

$$f_Y(\mathbf{y}) = f_X(g^{-1}(\mathbf{y}))\left|J_{g^{-1}}(\mathbf{y})\right|$$

where $J_{g^{-1}}$ is the Jacobian of g^{-1}.

E.g. let X be Y independent and set $(U, V) = (2X, 2X + Y)$.
Then $g^{-1}(u, v) = (u/2, v - u)$, so

$$|\mathbf{J}_{g^{-1}}| = \begin{vmatrix} 1/2 & 0 \\ -1 & 1 \end{vmatrix} = \frac{1}{2}$$

and $f_{U,V}(u, v) = f_X(u/2)f_Y(v - u)/2$.

Finite Mixtures (Optional)

Let $(F_1, ..., F_n)$ be a collection of CDFs with PDFs $(f_1, ..., f_n)$. Let $\mathbf{w} = (w_1, ..., w_n)$ a weight vector summing to 1.

A **finite mixture** is a random variable with CDF/PMF given by

$$F(x) = \sum_{i=1}^{n} w_i F_i(x), \quad f(x) = \sum_{i=1}^{n} w_i f_i(x).$$

Implementation: sample an index $i \in \{1, ..., n\}$ with probability w_i, then sample from f_i.

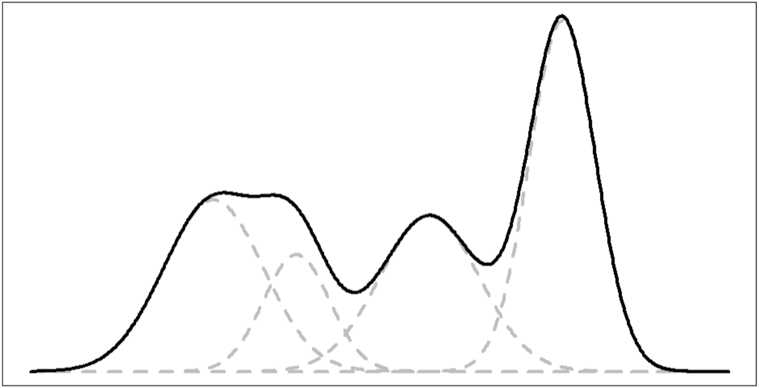
Extremely important: clustering, density estimation, image segmentation, zero inflated Poisson, etc.

Example (Optional)

Key example is a location-scale mixture of univariate normals:

$$f(x) = \sum_{i=1}^{n} w_i N(x; \mu_i, \sigma_i^2).$$

Note we could relabel the groups (not identifiable).



Other Techniques (Optional)

May need a more flexible distribution. Leads to **infinite mixture** models:

$$f(x) = \sum_{i=1}^{\infty} w_i f_i(x) \quad \text{or even} \quad f(x) = \int f(x|w)p(w)dw.$$

May need complicated multivariate distributions (e.g., correlated Poissons) with asymmetric correlation structures (e.g., stock crashes). Leads to **copulas.**

Exercises

1. Let $X_1, ..., X_n$ be iid with CDF F and PDF f. Order the variables $X_{(1)} \leq \cdots \leq X_{(n)}$.
   a) Find the CDF and PDF for $X_{(1)}$.
   b) Find the CDF and PDF for $X_{(n)}$.
   c) Evaluate these quantities when $X_i \sim Exp(\lambda)$.
2. Let $X \sim Gamma(\alpha, 1)$ and $Y \sim Gamma(\beta, 1)$ be independent.
   a) Write down the joint density, $f_{X,Y}(x, y)$.
   b) Find the joint density of $(U, V) = (X/(X+Y), X+Y)$.
   c) Identify the marginal distributions of U and V.

Solutions

1. a) For the min, work with the survival function:

$$P[X_{(1)} > x] = P[X_1 > x, ..., X_n > x] = \prod_{i=1}^{n}(1 - F(x))$$
$$= (1 - F(x))^n.$$

Thus $F_{X_{(1)}}(x) = 1 - (1 - F(x))^n$ and
$f_{X_{(1)}} = n(1 - F(x))^{n-1} f(x)$.

1. b) For the max, work with the cdf:

$$P[X_{(n)} \leq x] = P[X_1 \leq x, ..., X_n \leq x] = \prod_{i=1}^{n} F(x)$$
$$= F(x)^n.$$

Thus $F_{X_{(n)}}(x) = F(x)^n$ and $f_{X_{(n)}} = nF(x)^{n-1} f(x)$.

Solutions

1. c) Skipped - plug in $F_X(x) = 1 - \exp(-\lambda x)$ and simplify.
2. a) By independence, the joint density is the product:

$$\begin{aligned} f_{X,Y}(x,y) &= f_X(x) f_Y(y) \\ &= \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-x\} \frac{1}{\Gamma(\beta)} y^{\beta-1} \exp\{-y\} \end{aligned}$$

for $x, y > 0$.

Solutions

2. b) We need the inverse map to apply change of variables. Note

$$VU = (X+Y)\left(\frac{X}{X+Y}\right) = X.$$

This lets us solve for Y:

$$V - VU = (X+Y) - X = Y.$$

The inverse map is $(X, Y) = g^{-1}(U, V) = (VU, V(1-U))$ with Jacobian

$$J_{g^{-1}}(u, v) = \begin{pmatrix} \partial_u (g^{-1})_1 & \partial_v (g^{-1})_1 \\ \partial_u (g^{-1})_2 & \partial_v (g^{-1})_2 \end{pmatrix} = \begin{pmatrix} v & u \\ -v & 1-u \end{pmatrix}$$

The determinant is $|J_{g^{-1}}(u, v)| = v(1-u) + uv = v$. Note this is non-negative.

Solutions

2. b) By the change of variables formula,

$$\begin{aligned} f_{U,V}(u,v) &= \frac{1}{\Gamma(\alpha)}(vu)^{\alpha-1}\exp\{-vu\} \\ &\quad \times \frac{1}{\Gamma(\beta)}(v(1-u))^{\beta-1}\exp\{-v(1-u)\}v \\ &= \frac{1}{\Gamma(\alpha)\Gamma(\beta)}u^{\alpha-1}(1-u)^{\beta-1}v^{\alpha-1+\beta-1+1} \\ &\quad \times exp(-v+uv-uv) \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}u^{\alpha-1}(1-u)^{\beta-1}\frac{1}{\Gamma(\alpha+\beta)}v^{\alpha+\beta-1}\exp(-v). \end{aligned}$$

Here the supports are $u = x/(x+y) \in (0,1)$ and $v = x+y \in (0,\infty)$.

2. c) The PDF factors, hence U, V are *independent*. We read off $U \sim Beta(\alpha,\beta)$ and $V \sim Gamma(\alpha+\beta, 1)$.

Acknowledgements

Past contributors:

- Jordan Bryan, PhD student
- Brian Cozzi, MSS alumni
- Michael Valancius, MSS alumni
- Graham Tierney, PhD student
- Becky Tang, PhD student