QBS 120 - Lecture 3
Joint distributions
(Rice Chapter 3)

Rob Frost

Outline

- Joint distributions
- Independent RVs
- Conditional distributions
- Functions of jointly distributed RVs
- Extrema and order statistics

Joint distributions

- Nearly all scientific experiments capture multiple interrelated variables that can easy be modeled by a RV, e.g., treatment type and disease outcome, expression of multiple genes, etc.
- In this context, we are interested in the joint probability structure of the RVs.
- Joint CDF of RVs X and Y:

$$F_{X,Y}(x, y) = F(x, y) = P(X \leq x, Y \leq y)$$

If A is event that $X \leq x$ and B is event that $Y \leq y$, then $F(x, y) = P(A \cap B)$

Joint distributions, continued

Joint probability that X and Y belong to intervals $[x_1, x_2]$ and $[y_1, y_2]$:

$$P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) =$$
$$F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1)$$

Extends to more than 2 RVs:

$$F(x_1, x_2, ..., x_n) = P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n)$$

Discrete joint distributions

For discrete RVs, the joint PMF is defined as:

$$p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P(X_1 = x_1,\ldots,X_n = x_n)$$

Helpful to think of joint distribution in terms of a contingency table. For discrete RVs X and Y, the joint distribution can be represented by a 2-dimensional table (sum of entries = 1):

	y_1	$\ldots$	y_n
x_1	$p(x_1, y_1)$	$\ldots$	$p(x_1, y_n)$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
x_n	$p(x_n, y_1)$	$\ldots$	$p(x_n, y_n)$

Example: coin tossing

- Coin is tossed 3 times.
- $\Omega = \{hhh, hht, hth, htt, thh, tht, tth, ttt\}$
- Define two discrete RVs:
  - X: was the first toss a head? (Bernoulli with $p = 0.5$)
  - Y: total number of heads (Binomial with $p = 0.5, n = 3$)
- What is the joint distribution of X and Y?

Example: coin tossing, continued

Represent joint distribution using a contingency table.

	$Y = 0$	1	2	3
$X = 0$	1/8	1/4	1/8	0
$X = 1$	0	1/8	1/4	1/8

Y: total # of heads, X: heads on first toss?,
$\Omega = \{hhh, hht, hth, htt, thh, tht, tth, ttt\}$

Example: coin tossing, continued

Represent joint distribution using a contingency table.

	$Y=0$	1	2	3
$X=0$	1/8	1/4	1/8	0
$X=1$	0	1/8	1/4	1/8

Y: total # of heads, X: heads on first toss?,
$\Omega = \{hhh, hht, hth, htt, thh, tht, tth, ttt\}$

The rows and columns can be summed to give the marginal distributions of X and Y, i.e., the non-joint distributions:

	$Y=0$	1	2	3	
$X=0$	1/8	1/4	1/8	0	1/2
$X=1$	0	1/8	1/4	1/8	1/2
	1/8	3/8	3/8	1/8	1

Multinomial distribution

- n independent trials that can each result in one of r outcomes with probabilities $p_1, \ldots, p_r$, $\sum_{i=1}^{r} p_i = 1$.
- A discrete joint generalization of the binomial distribution (if r=2, have a binomial RV).
- The results of the trials can be represented by the vector $N_1, \ldots, N_r$ where N_i is the total number of outcomes of type i.
- Such results are often represented graphically as a histogram.
- Interested in $p_{N_1,\ldots,N_r}(n1, \ldots, n_r)$ (a discrete joint PMF)

Blood type example

- Assume we represent 4 distinct blood types: A, B, AB and O
- The number of outcomes/groups $r = 4$ and each will have an associated probability: (p_A, p_B, p_{AB}, p_O)
- The result of a specific experiment is the vector (n_A, n_B, n_{AB}, n_O).
- Many common biomedical variables are multinomial: ethnicity, cancer stage, etc.

Multinomial distribution, continued

How to compute $p_{N_1,\dots,N_r}(n_1,\dots,n_r)$?

Multinomial distribution, continued

How to compute $p_{N_1,\ldots,N_r}(n_1,\ldots,n_r)$?

- Because each trial is independent, the probability of any specific sequence of trials with those event counts is $p_1^{n_1}\ldots p_r^{n_r}$

Multinomial distribution, continued

How to compute $p_{N_1,\ldots,N_r}(n_1,\ldots,n_r)$?

- Because each trial is independent, the probability of any specific sequence of trials with those event counts is $p_1^{n_1}\ldots p_r^{n_r}$
- Since there are $n!/(n_1!\ldots n_r!)$ such sequences of n trials with those event counts, the joint PMF is:

$$\begin{aligned} p_{N_1,\ldots,N_r}(n_1,\ldots,n_r) &= P(N_1 = n_1,\ldots,N_r = n_r) \\ &= \binom{n}{n_1\ldots n_r} p_1^{n_1}\ldots p_r^{n_r} \end{aligned}$$

Blood type example

- Assume $P(A) = 0.4, P(B) = 0.11, P(AB) = 0.04, P(O) = 0.45$ for the population of interest.
- What is the probability that a random sample of 100 individuals has the composition: $N_A = 30, N_B = 20, N_{AB} = 6, N_O = 44$?

```
> dmultinom(x=c(30,20,6,44),
+           prob=c(0.4, 0.11, 0.04, 0.45))
[1] 1.323278e-05
```

Continuous joint distributions

If X and Y are continuous RVs, their joint density function is $f(x, y)$ with the following properties:

- Piecewise continuous function of X and Y.
- Nonnegative.
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx\ dy = 1$

CDF:

$$P(X \leq x, Y \leq y) = F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) du\ dv$$

Generalizes as expected for 3 or more RVs.

Continuous joint distributions, continued

Density from CDF:

$$f(x, y) = \frac{\delta^2}{\delta x \delta y} F(x, y)$$

Probability in small region near (x, y) is proportional to $f(x, y)$:

$$P(x \leq X \leq x + \delta_x, y \leq Y \leq y + \delta_y) \approx f(x, y)\delta_x \delta_y$$

Marginal CDF and density

Marginal CDF:

$$F_X(x) = P(X \le x) = \lim_{y \to \infty} F(x, y)$$
$$= \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u, y) dy \ dv$$

Marginal density:

$$f_X(x) = F'_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

Marginal discrete: sum joint PMF over other variable; marginal continuous: integrate density over other variable.

Bivariate normal distribution

For normal RVs X and Y, the joint density is:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} exp\Big(- \frac{1}{2(1-\rho^2)}\Big[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\Big]\Big)$$

Quite complex! How to interpret?

- μ_x, μ_y are the mean values of X and Y
- σ_x, σ_y are the standard deviations of X and Y
- ρ is the correlation between X and Y

Bivariate normal distribution, continued

First, find the marginal distributions of X and Y:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dx$$
$$= \frac{1}{\sigma_x\sqrt{2\pi}} e^{-0.5[(x-\mu_x)^2/\sigma_x^2]}$$

So, $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$. Similar result for Y:

$$f_Y(y) = \frac{1}{\sigma_y\sqrt{2\pi}} e^{-0.5[(y-\mu_y)^2/\sigma_y^2]}$$

Note that these marginal distributions do not include ρ.
What does ρ represent?

Bivariate normal distribution, continued

Reinvestigate $f(x, y)$ and note which terms are constants:

$$f(x, y) = C_1 exp\Big(- C_2 \Big[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x \sigma_y} \Big] \Big)$$

So, the density is constant whenever:

$$\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x \sigma_y} = \text{constant}$$

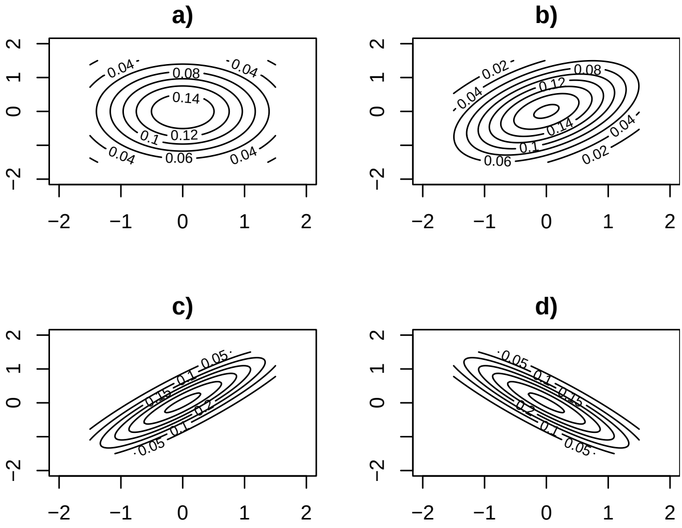
Recognize that as the equation of an ellipse in x and y. So, the density of the bivariate normal distribution has elliptical contours. The axes of the ellipse align with x and y axes whenever $\rho = 0$.

Bivariate normal distribution, continued

- The density of the bivariate normal distribution has elliptical contours of constant density.
- Maximum density at $x = \mu_x, y = \mu_y$
- σ_x and σ_y control the length of the axes.
- The axes of the ellipse align with x and y axes whenever $\rho = 0$. If $\rho \neq 0$, axes are tilted.

Bivariate normal distribution, continued

a) $\rho = 0$, b) $\rho = 0.5$, c) $\rho = 0.9$, d) $\rho = -0.9$,
$\mu_x = 0, \mu_y = 0, \sigma_x = 1, \sigma_y = 1$ in all panels.



Bivariate normal distribution, continued

R logic to generate bivariate normal density plots:

```
> library(mvtnorm)
> par(mfrow=c(2,2), mar=c(4,4,2,0))
> plotBivarNorm = function(mu = c(0,0), sigma, title) {
+         x.points = y.points = seq(-1.5,1.5,length.out=100)
+         z =matrix(0,nrow=100,ncol=100)
+         for (i in 1:100) {
+                 for (j in 1:100) {
+                         z[i,j] =dmvnorm(c(x.points[i],y.points[j]),
+                                         mean=mu,sigma=sigma)
+                 }
+         }
+         contour(x.points, y.points, z, xlim=c(-2,2), ylim=c(-2,2), main=title)
+ }
> plotBivarNorm(sigma = matrix(c(1,0,0,1), nrow=2, byrow=T), title="a)")
> plotBivarNorm(sigma = matrix(c(1,.5,.5,1), nrow=2, byrow=T), title="b)")
> plotBivarNorm(sigma = matrix(c(1,.9,.9,1), nrow=2, byrow=T), title="c)")
> plotBivarNorm(sigma = matrix(c(1,-.9,-.9,1), nrow=2, byrow=T), title="d)")
```

Multivariate normal distribution

The joint distribution of $p \geq 2$ normal RVs is described by the multivariate normal distribution:

$$f(x_1, ..., x_p) = \frac{1}{\sqrt{(2\pi)^p |\boldsymbol{\Sigma}|}} e^{-0.5(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

where

- $\mathbf{x}$ is the vector $(x_1, ..., x_p)$
- $\boldsymbol{\mu}$ is a vector of variable means $(\mu_1, ..., \mu_p)$
- $\boldsymbol{\Sigma}$ is a $p \times p$ positive definite covariance matrix, with diagonals equal to variances (σ_i^2) and off-diagonals equal to covariances $(\rho_{i,j}\sigma_i\sigma_j)$. For bivariate case:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

Outline

- Joint distributions
- **Independent RVs**
- Conditional distributions
- Functions of jointly distributed RVs
- Extrema and order statistics

Independent RVs

- RVs $X_1, ..., X_p$ are independent if the joint CDF factors into the product of marginal CDFs:

$$F(x_1, ..., x_p) = F_{X_1}(x_1)...F_{X_p}(x_p)$$

- Holds for both discrete and continuous RVs.
- Factoring of the joint CDF implies factoring of the joint density (or PMF):

$$\begin{aligned} f(x_1, ..., x_p) &= \frac{\delta}{\delta x_1 ... \delta x_p} F_{X_1}(x_1)...F_{X_p}(x_p) \\ &= f_{X_1}(x_1)...f_{X_p}(x_p) \end{aligned}$$

- Functions of independent RVs are themselves independent.

Independent normal RVs

For normal RVs X and Y, their joint density is:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} exp\Big(- \frac{1}{2(1-\rho^2)}\Big[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\Big]\Big)$$

What if $\rho = 0$?

Independent normal RVs

For normal RVs X and Y, their joint density is:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} exp\Big(-\frac{1}{2(1-\rho^2)}\Big[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\Big]\Big)$$

What if $\rho = 0$?

$$\begin{aligned} f(x,y) &= \frac{1}{2\pi\sigma_x\sigma_y} exp\Big(-\frac{1}{2}\Big[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\Big]\Big) \\ &= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-(x-\mu_x)^2/2\sigma_x^2} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-(y-\mu_y)^2/2\sigma_y^2} \\ &= f_Y(y) f_X(x) \end{aligned}$$

Note: 0 correlation does not imply independence for all RVs

Outline

- Joint distributions
- Independent RVs
- **Conditional distributions**
- Functions of jointly distributed RVs
- Extrema and order statistics

Conditional distributions for discrete RVs

For discrete RVs X and Y, the conditional probability that $X = x$ given that $Y = y$ and $p(y) > 0$ follows from the definition of conditional probability:

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

This can be represented using conditional, joint and marginal PMFs:

$$p_{X|Y}(x|y) = \frac{p_{XY}(X = x, Y = y)}{p_Y(y)}$$

Note: if $P(Y = y) = 0$, conditional probability/frequency is 0.

Coin flipping, revisited

- Coin is tossed 3 times.
- Define two discrete RVs:
  - X: was the first toss a head? (Bernoulli with $p = 0.5$)
  - Y: total number of heads (Binomial with $p = 0.5, n = 3$)
- Represent joint distribution using a contingency table:

	$Y = 0$	1	2	3
$X = 0$	1/8	1/4	1/8	0
$X = 1$	0	1/8	1/4	1/8

- What is the probability of 1 head total if the first was toss was a head ($P(Y = 1|X = 1)$)?

Coin tossing, continued

$$P(Y = 1|X = 1) = \frac{P(Y = 1, X = 1)}{P(X = 1)}$$

The joint probability, $P(Y = 1, X = 1)$, is given by the cell value for $Y = 1$ and $X = 1$. The marginal probability, $P(X = 1)$, is given by the row margin for $X = 1$.

	$Y = 0$	1	2	3	
$X = 0$	1/8	1/4	1/8	0	0.5
$X = 1$	0	**1/8**	1/4	1/8	**0.5**
	1/8	3/8	3/8	1/8	1

So

$$P(Y = 1|X = 1) = \frac{1/8}{1/2} = 1/4$$

Marginal distributions from joint distributions

To understand why the separate distributions of X and Y are represented by the marginals, remember:

	$Y = 0$	1	2	3
$X = 0$	$P((Y = 0) \cap (X = 0))$	$\dots$	$\dots$	$\dots$
$X = 1$	$P((Y = 0) \cap (X = 1))$	$\dots$	$\dots$	$\dots$

Marginal distributions from joint distributions

To understand why the separate distributions of X and Y are represented by the marginals, remember:

	$Y=0$	1	2	3
$X=0$	$P((Y=0)\cap(X=0))$	$\dots$	$\dots$	$\dots$
$X=1$	$P((Y=0)\cap(X=1))$	$\dots$	$\dots$	$\dots$

Using defn of conditional probability:

	$Y=0$	1	2	3
$X=0$	$P(Y=0\|X=0)P(X=0)$	$\dots$	$\dots$	$\dots$
$X=1$	$P(Y=0\|X=1)P(X=1)$	$\dots$	$\dots$	$\dots$
	$\sum_{x\in\{0,1\}} P(Y=0\|X=x)P(X=x)$	$\dots$	$\dots$	$\dots$

Because $X=0$ and $X=1$ are disjoint and span Ω, by Law of Total Probability, $\sum_{x\in\{0,1\}} P(Y=0|X=x)P(X=x) = P(Y=0)$

Conditional distributions for continuous RVs

For continuous RVs X and Y, the conditional density of X given Y is:

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}, 0 < f_Y(y) < \infty$$

Similar to the discrete case, the Law of Total Probability can be used to find the marginal distribution of one RV from the conditional distribution and marginal distribution of another RV:

$$\begin{aligned} f_{XY}(x,y) &= f_{X|Y}(x|y) f_Y(y) \\ \int_{-\infty}^{\infty} f_{XY}(x,y) &= \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y) \\ f_X(x) &= \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y) \end{aligned}$$

Conditional distribution for bivariate normal RVs

For normal RVs X and Y, the conditional density of Y given X is:

$$f_{Y|X}(y|x) = \frac{1}{\sigma_Y\sqrt{2\pi(1-\rho^2)}} exp\Big(\frac{-0.5\big[y-\mu_Y-\rho\sigma_Y/\sigma_X(x-\mu_X)\big]^2}{\sigma_Y^2(1-\rho^2)}\Big)$$

As expected, when $\rho = 0$, the conditional density simplifies to the $f_Y(y)$:

$$\frac{1}{\sigma_Y\sqrt{2\pi}} exp\Big(\frac{-0.5\big[y-\mu_Y\big]^2}{\sigma_Y^2}\Big)$$

Rejection method (Example D, 3.5.2)

Challenge: How to simulate from density function $f(x)$ if the inverse CDF cannot be computed?

Approach: Rejection method

Informal process:

- Imagine sketching f(x) on a rectangular board.
- Throw darts randomly at the board.
- If a dart lands below f(x) on the board, the horizontal position of the dart is the simulated value.

Rejection method, continued

Formal description:

- Assume $f(x)$ is nonzero on $[a, b]$ where a, b can be infinite.
- Choose a function $M(x)$ where:
  - $M(x) \geq f(x), a \leq x \leq b$ (i.e., the board must fully enclose f(x))
  - $m(x) = M(x) / \int_a^b M(x)dx$ is a density function
  - Easy to simulate from $m(x)$
  - For finite a, b, can pick m as $U(a, b)$ (i.e., uniform density is rectangular)
- To simulate:
  - Generate T from density m (horizontal position of dart)
  - Generate $U \sim U(0, 1)$ and independent of T (i.e., vertical position of dart)
  - If $M(T)U \leq f(T)$, accept and set $X = T$, otherwise, reject and repeat (i.e., is dart above or below f(x) line)

Rejection method, example

Example: Generate standard normal RVs between $[-3, 3]$ with m as $U(-3, 3)$.

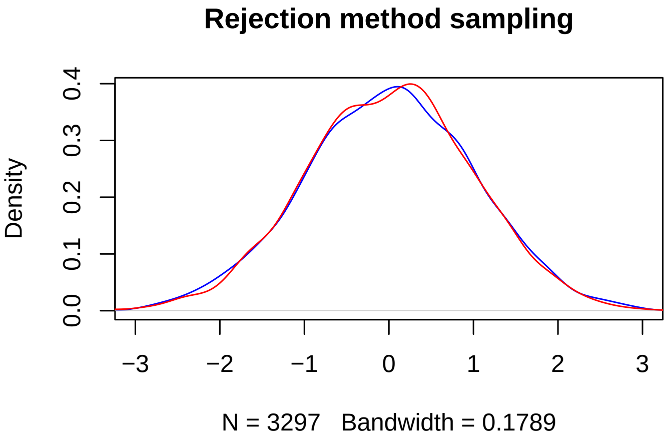
- $M(x) = 1/2$
- $m(x) = M(x)/\int_{-3}^{3} M(x)dx = 0.5/3 = 1/6$ (which is U(-3, 3))
- Is $M(x) \geq \phi(x), -3 \leq x \leq 3$?
  Yes: $1/2 > \phi(0) = 0.398942280401433$

```
> n=10000; min=-3; max=3
> t = runif(n, min=min, max=max)
> M_t = .5
> u = runif(n)
> accepts = which(M_t*u <= dnorm(t))
> length(accepts)/n
```

```
[1] 0.3297
```

Rejection method, example

```
> plot(density(t[accepts]), col="blue", xlim=c(min,max),
+         main="Rejection method sampling")
> lines(density(rnorm(length(accepts))), col="red")
```



Rejection method, example

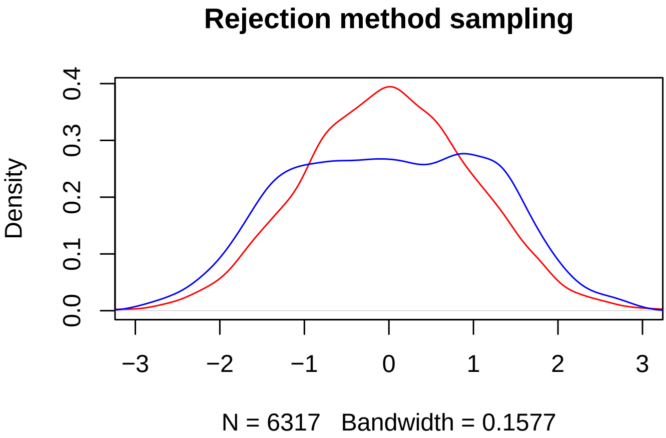
What if $M(x) \leq \phi(x)$? Pick $M(x) = 1/6$.

Note that $m(x) = M(x) / \int_{-3}^{3} M(x) dx$ still holds.

```
> length(accepts)/n
[1] 0.3297
> accepts = which((1/6)*u <= dnorm(t))
> length(accepts)/n
[1] 0.6317
```

Rejection method, example

```
> plot(density(rnorm(length(accepts))), col="red",, xlim=c(mi
+          main="Rejection method sampling")
> lines(density(t[accepts]), col="blue")
```



Rejection method, example

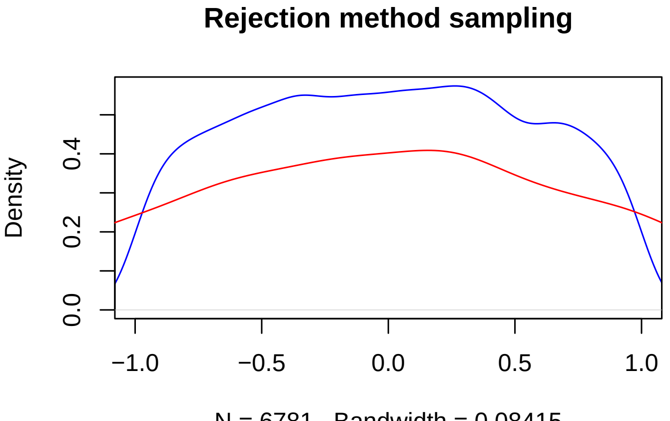
- Q: What approximation have I made to the rejection method?

Rejection method, example

- Q: What approximation have I made to the rejection method?
- A: I used $[-3, 3]$ rather than $[-\infty, \infty]$:
  $\int_{-3}^{3} \phi(x) dx \neq 1$ (but close,
  $\Phi(3) - \Phi(-3) =$0.99730020393674)
- What if $[-1, 1]$ was used instead? How well would simulated density match $\mathcal{N}(0, 1)$?

Rejection method, example

```
> min=-1; max=1; t = runif(n, min=min, max=max)
> accepts = which(M_t*u <= dnorm(t))
> plot(density(t[accepts]), col="blue", xlim=c(min,max),
+         main="Rejection method sampling")
> lines(density(rnorm(length(accepts))), col="red")
```



Rejection method, continued

For general case, show that generated density is $f(x)$:

$$
\begin{aligned}
P(x \le X \le x + dx) &= P(X \in dx) = P(T \in dx | accept) \\
&= \frac{P(T \in dx \cap accept)}{P(accept)} && \text{by defn. cond. prob} \\
&= \frac{P(accept | T \in dx) P(T \in dx)}{P(accept)} && \text{by defn. cond. prob} \\
&= \frac{P(U \le f(x)/M(x)) P(T \in dx)}{P(accept)} && \text{by rej. meth.} \\
&= \frac{(f(x)/M(x)) P(T \in dx)}{P(accept)} && \text{unif. CDF} \\
&= \frac{(f(x)/M(x)) m(x) dx}{P(accept)} && \text{by defn.} \\
&= \frac{(f(x)dx/M(x))(M(x)/\int_a^b M(x))dx}{P(accept)} && \text{by defn. of m(x)} \\
&= \frac{f(x)dx/\int_a^b M(x)dx}{P(accept)} && \text{simplify}
\end{aligned}
$$

Rejection method, continued

$$
\begin{aligned}
P(x \le X \le x + dx) &= \frac{f(x)dx / \int_a^b M(x)dx}{P(accept)} \\
&= \frac{f(x)dx / \int_a^b M(x)dx}{P(U \le f(T)/M(T))} && \text{by rej. meth.} \\
&= \frac{f(x)dx / \int_a^b M(x)dx}{\int_a^b f(t)m(t)/M(t)dt} && \text{law of total prob} \\
&= \frac{f(x)dx / \int_a^b M(x)dx}{\int_a^b (f(t)M(t))/(M(t)\int_a^b M(t)dt)dt} && \text{defn. of m(t)} \\
&= \frac{f(x)dx / \int_a^b M(x)dx}{1/\int_a^b M(t)dt} && \text{integration} \\
&= f(x)dx && \square
\end{aligned}
$$

Outline

- Joint distributions
- Independent RVs
- Conditional distributions
- **Functions of jointly distributed RVs**
- Extrema and order statistics

Functions of jointly distributed RVs

Following the development in Section 3.6 of Rice, we'll cover three cases:

- Sums: $Z = X + Y$
- Quotients: $Z = X/Y$
- General: $u = g_1(x, y), v = g_2(x, y)$

Sum of discrete RVs

Let X and Y be discrete random variables with joint PMF $p_{XY}(x, y)$. Define RV Z as $Z = X + Y$, what is the distribution of Z?

- Note that $Z = z$ when $X = x$ and $Y = z - x$
- So $p_Z(z)$ can be defined by the sum of $p_{XY}(x, y)$ for all possible values of $(x, z - x)$:

$$p_Z(z) = \sum_{x=-\infty}^{\infty} p_{XY}(x, z - x)$$

- If X and Y are independent (convolution of p_X, p_Y):

$$p_Z(z) = \sum_{x=-\infty}^{\infty} p_X(x) p_Y(z - x)$$

Sum of continuous RVs

Let X and Y are continuous RVs with joint density $f_{XY}(x, y)$. Approach is similar but start with CDF and use $y = z - x$ as an integration limit.

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f_{XY}(x, y) dy dx$$

Change $y = v - x$, switch order of integration and take δ/δ_z:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(x, z - x) dx$$

If X and Y independent, get the convolution

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx$$

Lifetime of component with backup (Rice 3.6.1 Example A)

Assume the lifetime of a given component (T_1) has an exponential distribution (parameter λ) and there is an independent and identical backup (T_2). Total system lifetime is therefore a sum of the component lifetimes: $S = T_1 + T_2$.

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx \\
f_S(s) &= \int_0^s f_{T_1}(t) f_{T_2}(s-t)\,dt && \text{note limits} \\
&= \int_0^s \lambda e^{-\lambda t} \lambda e^{-\lambda(s-t)}\,dt && \text{plug-in exp densities} \\
&= \lambda^2 \int_0^s e^{-\lambda s}\,dt && \text{simplify} \\
&= \lambda^2 s e^{-\lambda s} && \text{a gamma density}
\end{aligned}
$$

Quotient of RVs

For $Z = Y/X$, similar logic and integration manipulations leads to:

$$f_Z(z) = \int_{-\infty}^{\infty} |x| f_{XY}(x, xz) dx$$

For independent X and Y, this factors as:

$$f_Z(z) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(xz) dx$$

Example: Cauchy distribution is ratio of independent $\mathcal{N}(0, 1)$ RVs

General functions of joint RVs

Let X and Y be jointly distributed continuous RVs. Define RVs U and V via the functions:

$$u = g_1(x, y), v = g_2(x, y)$$

and inverse transformations:

$$x = h_1(u, v), y = h_2(u, v)$$

Define the Jacobian of X and Y as:

$$J(x, y) = det \begin{bmatrix} \frac{\delta g_1}{\delta x} & \frac{\delta g_1}{\delta y} \\ \frac{\delta g_2}{\delta x} & \frac{\delta g_2}{\delta y} \end{bmatrix} = (\frac{\delta g_1}{\delta x})(\frac{\delta g_2}{\delta y}) - (\frac{\delta g_2}{\delta x})(\frac{\delta g_1}{\delta y})$$

Joint density of U and V is:

$$f_{UV}(u, v) = f_{XY}(h_1(u, v), h_2(u, v))|J^{-1}(h_1(u, v), h_2(u, v))|$$

Linear functions of normal RVs (Example C in Rice 3.6.2)

Let X_1 and X_2 be independent standard normal RVs.
Define:

$$y_1 = g_1(x_1, x_2) = x_1$$
$$y_2 = g_2(x_1, x_2) = x_1 + x_2$$

- $J(x, y) = det \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = 1$
- $x_1 = h_1(y_1, y_2) = y_1$, $x_2 = h_2(y_1, y_2) = y_2 - y_1$,

What is $f_{Y_1 Y_2}(y_1, y_2)$?

Linear functions of normal RVs, continued

By definition:

$$f_{Y_1Y_2}(y_1, y_2) = f_{X_1X_2}(h_1(y_1, y_2), h_2(y_1, y_2))|J^{-1}(h_1(y_1, y_2), h_2(y_1, y_2))|$$

Simplify

$$\begin{aligned}
f_{Y_1Y_2}(y_1, y_2) &= f_{X_1X_2}(h_1(y_1, y_2), h_2(y_1, y_2)) && \text{Jacobian is 1} \\
&= f_{X_1}((h_1(y_1, y_2))f_{X_2}(h_2(y_1, y_2)) && X_1 \text{ and } X_2 \text{ are independent} \\
&= f_{X_1}(y_1)f_{X_2}(y_2 - y_1) && \text{sub in } h_1 \text{ and } h_2 \\
&= \frac{1}{\sqrt{2\pi}}e^{-y_1^2/2}\frac{1}{\sqrt{2\pi}}e^{-(y_2-y_1)^2/2} && X_1, X_2 \sim \mathcal{N}(0, 1) \\
&= \frac{1}{2\pi}e^{-1/2(2y_1^2+y_2^2-2y_1y_2)} && \text{bivariate normal distribution}
\end{aligned}$$

$\rightarrow$ A linear transformation of bivariate normal RVs is also bivariate normal.

Outline

- Joint distributions
- Independent RVs
- Conditional distributions
- Functions of jointly distributed RVs
- **Extrema and order statistics**

Extrema and order statistics

- Let $X_1, \ldots, X_n$ be independent and identically distributed RVs with CDF $F_X()$ and density $f_X()$.
- Let U be the maximum of the X_i and let V be the minimum of the X_i (extrema of the X_i)
- More generally, the X_i can be sorted: $X_{(1)} \leq \ldots \leq X_{(n)}$. $X_{(k)}$ is considered the kth-order statistic (U is the nth-order statistic and V is the 1st-order statistic)

What are the distributions of U, V and $X_{(k)}$?

Extrema distributions

Note that $U \leq u$ only if all $X_i \leq u$ (it is the maximum). So:

$$
\begin{aligned}
F_U(u) &= P(U \leq u) \\
&= P(X_1 \leq u)P(X_2 \leq u)...P(X_n \leq u) \\
&= [F_X(u)]^n \\
f_U(u) &= \frac{\delta}{\delta u}[F_X(u)]^n \\
&= nf_X(u)[F_X(u)]^{n-1}
\end{aligned}
$$

Similar approach can be used to find $F_V()$ and $f_V()$:

$$
\begin{aligned}
F_V(v) &= 1 - [1 - F_X(v)]^n \\
f_V(v) &= nf_X(v)[1 - F_X(v)]^{n-1}
\end{aligned}
$$

Example: components in series

Suppose n components are connected in series and each has an exponential lifetime with parameter λ. What is the lifetime of the entire system?

Since the system will fail if any single component fails, the system lifetime is the minimum life of any single component.

$$
\begin{aligned}
f_V(v) &= n f_X(v)[1 - F_X(v)]^{n-1} \\
&= n\lambda e^{-\lambda v}(e^{-\lambda v})^{n-1} \\
&= n\lambda e^{-n\lambda v}
\end{aligned}
$$

So, exponential distribution with parameter $n\lambda$

Example: components in series, continued

If $\lambda = 0.1$, how do the system lifetime densities compare for $n = 5$ and $n = 10$?

```
> t = seq(from=0, to=2, by=0.01)
> x = dexp(t, rate=.5)
> s = dexp(t, rate=1.0)
> plot(t,s, type="l", lty="dashed")
> points(t,x,type="l", lty="solid")
```