STAT 431 — Applied Bayesian Analysis — Course Notes

Probability Review

Fall 2022

Sample Space and Events

sample space: all possible outcomes (fully specified)

event: subset of sample space

Eg: Two coin flips

$$\Omega \;=\; \text{sample space} \;=\; \{HH, HT, TH, TT\}$$

$$\text{event } A \;=\; \text{both flips same} \;=\; \{HH, TT\}$$

[Illustrate ...]

The usual set operations apply to events:

$$
\begin{aligned}
A \cup B &= \textbf{union} \text{ of } A \text{ and } B \\
&= \text{outcomes in either (or both)} \\[1em]
A \cap B &= \textbf{intersection} \text{ of } A \text{ and } B \\
&= \text{outcomes in both} \\[1em]
\overline{A} &= \textbf{complement} \text{ of } A \\
&= \text{outcomes not in } A
\end{aligned}
$$

Events A and B are **disjoint** if

$$A \cap B = \varnothing \quad \text{(the null set)}$$

Probability

probability: to each event A, assigns a number $\mathrm{Prob}(A)$, with such properties as

- $0 \leq \mathrm{Prob}(A) \leq 1$
- $\mathrm{Prob}(\varnothing) = 0$ ($\varnothing$ is the "null event")
- $\mathrm{Prob}(\Omega) = 1$ (Ω is the sample space)
- if A and B are disjoint,

$$\mathrm{Prob}(A \cup B) = \mathrm{Prob}(A) + \mathrm{Prob}(B)$$

- if $A_1, \ldots, A_n$ are disjoint,

$$\text{Prob}\left(\bigcup_{i=1}^{n} A_i\right) \;=\; \sum_{i=1}^{n} \text{Prob}(A_i)$$

- $\text{Prob}(\overline{A}) \;=\; 1 \;-\; \text{Prob}(A)$

Eg: two fair coin flips

$$A = \text{both heads} \qquad B = \text{both tails}$$

$$\mathrm{Prob}(A) = ? \qquad \mathrm{Prob}(B) = ?$$

$$\mathrm{Prob}\left(\overline{A}\right) = ?$$

$$\mathrm{Prob}(A \cup B) = ? \qquad \mathrm{Prob}(A \cap B) = ?$$

Conditioning

If $\mathrm{Prob}(B) \neq 0$, the **conditional probability** of A given B is

$$\mathrm{Prob}(A \mid B) \;\;=\;\; \frac{\mathrm{Prob}(A \cap B)}{\mathrm{Prob}(B)}$$

($\mathrm{Prob}(A)$ is sometimes called the **marginal probability** of A)

Conditional probabilities behave like ordinary probabilities:

e.g. $\;0 \;\leq\; \mathrm{Prob}(A \mid B) \;\leq\; 1$

e.g. $\;\mathrm{Prob}\big(\overline{A} \mid B\big) \;=\; 1 \,-\, \mathrm{Prob}(A \mid B)$

Note:

$$
\begin{aligned}
\mathrm{Prob}(A \cap B) &= \mathrm{Prob}(A \mid B)\ \mathrm{Prob}(B) \qquad (\mathrm{Prob}(B) \neq 0) \\
&= \mathrm{Prob}(B \mid A)\ \mathrm{Prob}(A) \qquad (\mathrm{Prob}(A) \neq 0)
\end{aligned}
$$

In general,

$$\text{joint} \;=\; \text{conditional} \times \text{marginal}$$

Eg: two fair coin flips

$$A = \{HH, TT\} \qquad B = \{HH, HT, TH\}$$

$$\mathrm{Prob}(A) = ? \qquad \mathrm{Prob}(B) = ? \qquad \mathrm{Prob}(A \cap B) = ?$$

$$\mathrm{Prob}(B \mid A) = ?$$

$$\mathrm{Prob}(A \mid B) = ?$$

[Interpret ...]

Bayesian perspective:

$$
\begin{aligned}
M &= \text{a particular model for the data} \\
D &= \text{(event of) the data}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Prob}(D \mid M) &= \text{probability given to } D \text{ when } M \text{ is true} \\
&= \text{“likelihood”}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Prob}(M) &= \text{probability of } M \text{ with no other information} \\
&= \text{“prior”}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Prob}(M \mid D) &= \text{probability of } M \text{ after observing } D \\
&= \text{“posterior”}
\end{aligned}
$$

Bayes' Rule

[Illustrate sample space ...]

Notice:

$$\mathrm{Prob}(A) \;\; = \;\; \mathrm{Prob}\big((A \cap B) \cup (A \cap \overline{B})\big)$$

Bayes' Rule

[Illustrate sample space ...]

Notice:

$$\begin{aligned}\mathrm{Prob}(A) &= \mathrm{Prob}\big((A \cap B) \cup (A \cap \overline{B})\big) \\ &= \mathrm{Prob}(A \cap B) + \mathrm{Prob}\big(A \cap \overline{B}\big)\end{aligned}$$

Bayes' Rule

[Illustrate sample space ...]

Notice:

$$
\begin{aligned}
\mathrm{Prob}(A) &= \mathrm{Prob}\big((A \cap B) \cup (A \cap \overline{B})\big) \\
&= \mathrm{Prob}(A \cap B) + \mathrm{Prob}\big(A \cap \overline{B}\big) \\
&= \mathrm{Prob}(A \mid B)\ \mathrm{Prob}(B) + \mathrm{Prob}\big(A \mid \overline{B}\big)\ \mathrm{Prob}\big(\overline{B}\big)
\end{aligned}
$$

(provided $0 < \mathrm{Prob}(B) < 1$)

Bayes' rule (simple form):

$$\mathrm{Prob}(B \mid A) \quad = \quad \frac{\mathrm{Prob}(A \cap B)}{\mathrm{Prob}(A)}$$

Bayes' rule (simple form):

$$\mathrm{Prob}(B \mid A) \;=\; \frac{\mathrm{Prob}(A \cap B)}{\mathrm{Prob}(A)}$$

$$= \frac{\mathrm{Prob}(A \mid B)\ \mathrm{Prob}(B)}{\mathrm{Prob}(A \mid B)\ \mathrm{Prob}(B) \;+\; \mathrm{Prob}\big(A \mid \overline{B}\big)\ \mathrm{Prob}\big(\overline{B}\big)}$$

(provided all conditional probabilities exist)

Eg: Say the pop. of Cyprus is 80% Greek, 20% Turkish.

Suppose English is spoken by 90% of the Greek population and 50% of the Turkish population.

What's the prob. an English-speaking Cypriot is Greek?

$$A \;=\; \text{speaks English} \qquad\qquad B \;=\; \text{is Greek}$$

Expression for the desired probability?

$$\mathrm{Prob}(B) \;=\; ? \qquad\qquad \mathrm{Prob}\big(\overline{B}\big) \;=\; ?$$

$$\mathrm{Prob}(A \mid B) \;=\; ? \qquad\qquad \mathrm{Prob}\big(A \mid \overline{B}\big) \;=\; ?$$

Answer?

Now generalize ...

Suppose $B_1, B_2, B_3, \ldots$ form a **partition** of Ω:

- all are disjoint
- $\bigcup_{\text{all } j} B_j \;=\; \Omega$ (exhaustive)

Also, assume $\mathrm{Prob}(B_j) \neq 0$, all j.

Law of total probability:

$$\mathrm{Prob}(A) \;=\; \sum_{\text{all } j} \mathrm{Prob}(A \mid B_j)\ \mathrm{Prob}(B_j)$$

[Illustrate ...]

Bayes' rule (for probabilities):

If $B_1, B_2, \ldots$ is a partition,

$$\mathrm{Prob}(B_i \mid A) \quad = \quad \frac{\mathrm{Prob}(A \mid B_i)\ \mathrm{Prob}(B_i)}{\displaystyle\sum_{\text{all } j} \mathrm{Prob}(A \mid B_j)\ \mathrm{Prob}(B_j)}$$

(The previous special case had $B_1 = B$, $B_2 = \overline{B}$.)

Bayes' rule (for probabilities):

If $B_1, B_2, \ldots$ is a partition,

$$\mathrm{Prob}(B_i \mid A) \quad = \quad \frac{\mathrm{Prob}(A \mid B_i)\ \mathrm{Prob}(B_i)}{\sum_{\text{all } j} \mathrm{Prob}(A \mid B_j)\ \mathrm{Prob}(B_j)}$$

(The previous special case had $B_1 = B$, $B_2 = \overline{B}$.)

So

$$\mathrm{Prob}(B_i \mid A) \quad \propto \quad \mathrm{Prob}(A \mid B_i)\ \mathrm{Prob}(B_i)$$

(since the denominator doesn't depend on i)

Bayesian application:

$$\begin{aligned} M_1, M_2, \ldots &= \text{distinct models for the data} \\ D &= \text{(event of) the data} \end{aligned}$$

By Bayes' rule,

$$\begin{aligned} \mathrm{Prob}(M_i \mid D) &\propto \mathrm{Prob}(D \mid M_i)\, \mathrm{Prob}(M_i) \\ \text{posterior} &\propto \text{likelihood} \times \text{prior} \end{aligned}$$

The proportionality constant

$$1 \Bigg/ \sum_{\text{all } j} \mathrm{Prob}(D \mid M_j)\, \mathrm{Prob}(M_j)$$

is called the **normalizing constant** (in BSM).

Eg: COVID testing (revisited)

Independent Events

Events A and B are **independent** if

$$\mathrm{Prob}(A \cap B) \;=\; \mathrm{Prob}(A)\,\mathrm{Prob}(B)$$

(otherwise **dependent**)

If $\mathrm{Prob}(B) \neq 0$, this is the same as

$$\mathrm{Prob}(A \mid B) \;=\; \mathrm{Prob}(A)$$

$$(\text{conditional} \;=\; \text{marginal})$$

Events A and B are **conditionally independent** given C if

$$\mathrm{Prob}(A \cap B \mid C) \;\; = \;\; \mathrm{Prob}(A \mid C)\, \mathrm{Prob}(B \mid C)$$

Note: A and B are not necessarily independent!

Often there are data events that are independent conditional on the model:

$$\mathrm{Prob}(D_1 \cap D_2 \mid M) \;\; = \;\; \mathrm{Prob}(D_1 \mid M)\, \mathrm{Prob}(D_2 \mid M)$$

That is, the "likelihood" may factor.

Random Variables and Distributions

random variable: real-valued function on the sample space

May be ...

- **discrete**: takes values in a countable set

  e.g. binomial, geometric, Poisson

- **continuous**: takes values on a continuum

  e.g. normal, exponential, gamma

(For review: BSM, Appendix A.1)

The **distribution** of a random variable X is characterized by its **density**:

- Discrete density:

$$f(x) \;\;=\;\; \mathrm{Prob}(X = x)$$

  (sometimes called a PMF)

- Continuous density: $f(x)$ such that

$$\int_G f(x)\,dx \;\;=\;\; \mathrm{Prob}(X \in G)$$

  (often called a PDF)

The **joint distribution** of random variables X and Y can often be characterized by a **joint density**

$$f(x, y)$$

- Both discrete:

$$\begin{aligned} f(x, y) &= \mathrm{Prob}(X = x, Y = y) \\ &= \mathrm{Prob}\big(\{X = x\} \cap \{Y = y\}\big) \end{aligned}$$

- Jointly continuous:

$$\int_{G_1} \int_{G_2} f(x, y)\, dy\, dx = \mathrm{Prob}(X \in G_1, Y \in G_2)$$

The individual densities of X and Y are their **marginal densities**, which define their **marginal distributions**.

Eg:

$$f(x) = \begin{cases} \sum_{\text{all } y} f(x,y), & Y \text{ discrete} \\ \int f(x,y)\,dy, & Y \text{ continuous} \end{cases}$$

Conditioning

The **conditional distribution** of X given Y is characterized by the **conditional density**

$$f(x \mid y) \;=\; \frac{f(x,y)}{f(y)} \qquad \text{(wherever } f(y) > 0\text{)}$$

Conditioning

The **conditional distribution** of X given Y is characterized by the **conditional density**

$$f(x \mid y) \;=\; \frac{f(x,y)}{f(y)} \qquad \text{(wherever } f(y) > 0\text{)}$$

Note:

$$f(x,y) \;=\; f(x \mid y)\, f(y) \;=\; f(y \mid x)\, f(x)$$

are examples of the general form

$$\text{joint} \;=\; \text{conditional} \times \text{marginal}$$

This idea can be used to define the joint density when X and Y are of different types.

For example, if X is continuous and Y is discrete, let

$$f(x,y) \;=\; f(x \mid y)\, f(y) \;=\; f(y \mid x)\, f(x)$$

where

$$\begin{aligned} f(x \mid y) &= \text{a continuous density for each } y \\ f(y \mid x) &= \text{a discrete density for each } x \end{aligned}$$

(Use whichever of these is most convenient.)

A general process for working with the joint distribution of X and Y:

1. Specify the conditional density of Y given X
2. Specify the marginal density of X
3. Use the product of these densities as their joint density

Example: Uniform-Binomial

$$\begin{aligned} Y \mid X = x &\sim \text{Binomial}(n, x) \\ X &\sim \text{Uniform}(0, 1) \end{aligned}$$

(n is a given "number of trials", x is "success prob.")

Example: Uniform-Binomial

$$\begin{aligned} Y \mid X = x &\sim \text{Binomial}(n, x) \\ X &\sim \text{Uniform}(0, 1) \end{aligned}$$

(n is a given "number of trials", x is "success prob.")

So

$$f(y \mid x) = \binom{n}{y} x^y (1-x)^{n-y} \qquad y = 0, \ldots, n$$

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

... and the "joint density" is

$$f(y \mid x)\, f(x) \quad =$$

$$\begin{cases} \dbinom{n}{y} x^y (1-x)^{n-y} \;\cdot\; 1 & 0 < x < 1,\ y = 0, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

The marginal density for X is (of course)

$$f(x) \quad = \quad \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

The marginal density for Y is (for $y = 0, \ldots, n$)

$$
\begin{aligned}
f(y) &= \int f(y \mid x)\, f(x)\, dx \\
&= \int_0^1 \binom{n}{y} x^y (1-x)^{n-y} \cdot 1\, dx \\
&= \binom{n}{y} \int_0^1 \underbrace{x^y (1-x)^{n-y}}\, dx
\end{aligned}
$$

The marginal density for Y is (for $y = 0, \ldots, n$)

$$\begin{aligned} f(y) &= \int f(y \mid x)\, f(x)\, dx \\ &= \int_0^1 \binom{n}{y} x^y (1-x)^{n-y} \ \cdot\ 1\, dx \\ &= \binom{n}{y} \int_0^1 \underbrace{x^y (1-x)^{n-y}}_{\text{“kernel” of a Beta density}} \, dx \end{aligned}$$

The marginal density for Y is (for $y = 0, \ldots, n$)

$$
\begin{aligned}
f(y) &= \int f(y \mid x)\, f(x)\, dx \\
&= \int_0^1 \binom{n}{y} x^y (1-x)^{n-y} \cdot 1\, dx \\
&= \binom{n}{y} \int_0^1 \underbrace{x^y (1-x)^{n-y}}_{\text{"kernel" of a Beta density}} dx
\end{aligned}
$$

Density of $\mathrm{Beta}(\alpha, \beta)$ distribution ($\alpha, \beta > 0$):

$$
\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1} (1-x)^{\beta-1} \qquad 0 < x < 1
$$

(see BSM, Appendix A.1)

Thus, for $y = 0, \ldots, n$,

$$\begin{aligned} f(y) &= \binom{n}{y} \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)} \\ &\quad \cdot \int_0^1 \underbrace{\frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}\, x^y \, (1-x)^{n-y}}_{\mathrm{Beta}(y+1,\, n-y+1) \text{ density}} \, dx \end{aligned}$$

Thus, for $y = 0, \ldots, n$,

$$f(y) = \binom{n}{y} \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)}$$

$$\cdot \int_0^1 \underbrace{\frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}\, x^y\,(1-x)^{n-y}}_{\mathrm{Beta}(y+1,\, n-y+1) \text{ density}}\, dx$$

$$= \binom{n}{y} \frac{y!\,(n-y)!}{(n+1)!} \cdot 1$$

Thus, for $y = 0, \dots, n$,

$$
\begin{aligned}
f(y) &= \binom{n}{y} \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)} \\
&\qquad \cdot \int_0^1 \underbrace{\frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}\, x^y\, (1-x)^{n-y}}_{\mathrm{Beta}(y+1,\, n-y+1) \text{ density}}\, dx \\
&= \binom{n}{y} \frac{y!\,(n-y)!}{(n+1)!} \cdot 1 \\
&= \frac{n!}{y!\,(n-y)!} \frac{y!\,(n-y)!}{(n+1)\cdot n!} = \frac{1}{n+1}
\end{aligned}
$$

So the marginal distribution of Y is a "discrete uniform" distribution:

$$f(y) \quad = \quad \begin{cases} \dfrac{1}{n+1} & y = 0, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

[Illustrate ...]

Bayes' Rule

Bayes' rule (for densities):

$$f(x \mid y) \;=\; \frac{f(y \mid x)\, f(x)}{m}$$

where

$$m \;=\; f(y) \;=\; \begin{cases} \displaystyle\sum_{\text{all } x} f(y \mid x)\, f(x), & X \text{ discrete} \\[2ex] \displaystyle\int f(y \mid x)\, f(x)\, dx, & X \text{ continuous} \end{cases}$$

BSM would call $1/m$ the **normalizing constant**.

Bayesian application:

Suppose the model is (fully) defined by a **parameter** θ.

Let y be the observed data.

Then

$$\begin{array}{ccccc} f(\theta \mid y) & \propto & f(y \mid \theta) & \cdot & f(\theta) \\ \text{posterior} & \propto & \text{likelihood} & \times & \text{prior} \end{array}$$

where the proportionality is in θ (for fixed y).

More Variables

Densities can be extended to three or more variables.

E.g. X, Y, and Z could have a joint density defined by

$$f(x, y, z) \;\; = \;\; f(z \mid x, y)\; f(y \mid x)\; f(x)$$

where conditioning on two variables is defined as, e.g.

$$f(z \mid x, y) \;\; = \;\; \frac{f(x, y, z)}{f(x, y)} \qquad \text{(wherever } f(x, y) > 0\text{)}$$

The marginal densities would be denoted

$$f(x, y), \qquad f(x, z), \qquad f(y, z),$$

$$f(x), \qquad f(y), \qquad f(z)$$

Marginal densities are obtained by summing/integrating out the other variables, e.g.

$$f(x, z) = \begin{cases} \displaystyle\sum_{\text{all } y} f(x, y, z), & Y \text{ discrete} \\ \displaystyle\int f(x, y, z)\, dy, & Y \text{ continuous} \end{cases}$$

Similarly, joint conditionals can be defined as, e.g.

$$f(x, y \mid z) \;\; = \;\; \frac{f(x, y, z)}{f(z)} \qquad \text{(wherever } f(z) > 0\text{)}$$

Certain rules for marginal densities extend to conditional densities, e.g.

$$f(x, y \mid z) \;\; = \;\; f(y \mid x, z)\, f(x \mid z)$$

Independent Random Variables

X and Y are **independent** when they have a joint density that factors into marginals:

$$f(x, y) = f(x)\, f(y)$$

Note: If X and Y are independent,

$$f(x \mid y) = f(x) \qquad\qquad f(y \mid x) = f(y)$$

Note: If $f(x \mid y)$ doesn't depend on y (or if $f(y \mid x)$ doesn't depend on x), then X and Y are independent. (Why?)

Let Z be another random variable.

X and Y are **conditionally independent given** $Z = z$ if

$$f(x, y \mid z) \;\; = \;\; f(x \mid z)\, f(y \mid z)$$

In general, this does not imply that X and Y are (marginally) independent.

X and Y are **conditionally independent given** Z if

$$f(x, y \mid z) \;=\; f(x \mid z)\, f(y \mid z) \qquad \text{for all } z \quad (f(z) > 0)$$

This is (almost) equivalent to

$$f(x \mid y, z) \;=\; f(x \mid z)$$

and to

$$f(y \mid x, z) \;=\; f(y \mid z)$$

Features of Distributions

The **expected value** or **mean** of X is

$$\mathrm{E}(X) = \begin{cases} \displaystyle\sum_{\text{all } x} x\, f(x), & X \text{ discrete} \\ \displaystyle\int x\, f(x)\, dx, & X \text{ continuous} \end{cases}$$

A **median** m_X of X satisfies

$$\mathrm{Prob}(X < m_X) \leq 0.5 \quad \text{and} \quad \mathrm{Prob}(X > m_X) \leq 0.5$$

A **mode** of X is a value maximizing $f(x)$. It need not exist or be unique.

A τ-**quantile** $Q_X(\tau)$ of X satisfies

$$\mathrm{Prob}\big(X < Q_X(\tau)\big) \quad \leq \quad \tau$$

and

$$\mathrm{Prob}\big(X > Q_X(\tau)\big) \quad \leq \quad 1-\tau$$

If X is continuous,

$$\mathrm{Prob}\big(X \leq Q_X(\tau)\big) \quad = \quad \tau$$

[Illustrate ...]

The **variance** of X is

$$\mathrm{Var}(X) \;=\; \mathrm{E}\big((X - \mu_X)^2\big)$$

where $\mu_X = \mathrm{E}(X)$.

The **standard deviation** is just the square root of the variance:

$$\mathrm{SD}(X) \;=\; \sqrt{\mathrm{Var}(X)}$$

For two random variables, X and Y, their **covariance** is

$$\mathrm{Cov}(X,Y) \;\; = \;\; \mathrm{E}\big((X-\mu_X)(Y-\mu_Y)\big)$$

where $\mu_X = \mathrm{E}(X)$ and $\mu_Y = \mathrm{E}(Y)$.

Their **correlation** is

$$\rho_{XY} \;\; = \;\; \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\;\mathrm{Var}(Y)}}$$

The **conditional expected value** (or **conditional mean**) of X given $Y = y$ is

$$\mathrm{E}(X \mid Y = y) \;=\; \begin{cases} \displaystyle\sum_{\text{all } x} x\, f(x \mid y), & X \text{ discrete} \\ \displaystyle\int x\, f(x \mid y)\, dx, & X \text{ continuous} \end{cases}$$

The **conditional variance** of X given $Y = y$ is

$$\mathrm{Var}(X \mid Y = y) \;=\; \mathrm{E}\big((X - \mu_{X|y})^2 \mid Y = y\big)$$

where $\mu_{X|y} \;=\; \mathrm{E}(X \mid Y = y)$.

Similarly, we may be able to define conditional medians, conditional modes, conditional quantiles, etc.

Just replace any probability or density with an appropriate conditional probability or conditional density.

Notational note:

We sometimes write

$$\begin{aligned} \mathrm{E}(X \mid y) \quad &\text{for} \quad \mathrm{E}(X \mid Y = y) \\ \mathrm{Var}(X \mid y) \quad &\text{for} \quad \mathrm{Var}(X \mid Y = y) \end{aligned}$$

Similarly, write

$$X \mid y \ \sim \ \cdots \qquad \text{for} \qquad X \mid Y = y \ \sim \ \cdots$$

Change of Variables

Suppose X is continuous, with density $f(x)$, and let

$$Y = g(X)$$

where g has a differentiable inverse g^{-1}.

Then Y is continuous, with density

$$f(y) = f(x)\left|\frac{dx}{dy}\right|$$

where x is (implicitly) equal to $g^{-1}(y)$.

This **change of variables formula** can be more explicitly written as

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

where f_X and f_Y are densities of X and Y.