Conceptual Bootcamp

Kyle Burris and Abbas Zaidi

August 22, 2016

Agenda

- Probability Theory
  1. Change of Variables for Distributions
  2. Expectation of a Random Variable
  3. Iterated Expectation and Variance
  4. Bayes' Rule
  5. Central Limit Theorem
- Matrix Properties
- Useful Resources

Change of Variables

If Z has a density $f_Z(z)$ and $Y = G(Z)$, the density of Y is given as

$$f_Y(y) = f_Z(G^{-1}(y))|\det(dG^{-1})|$$

where dG^{-1} is the derivative (matrix of partial derivatives) of G^{-1} evaluated at y.

Change of Variables Example

Recall

$$f_Y(y) = f_Z(G^{-1}(y))|\det(dG^{-1})|$$

Suppose σ has probability density $f(\sigma) \propto 1/\sigma$. What is the probability density (up to proportionality) of $\log(\sigma)$?

$$\text{Let } G(\sigma) = \log(\sigma) \Rightarrow G^{-1}(\sigma) = e^{\sigma}$$

$$dG^{-1} = e^{\sigma}$$

$$f_{\sigma}(\sigma) \propto \frac{1}{e^{\sigma}}|e^{\sigma}| = 1$$

Change of Variables Example

Recall

$$f_Y(y) = f_Z(G^{-1}(y))|\det(dG^{-1})|$$

Suppose σ has probability density $f(\sigma) \propto 1/\sigma$. What is the probability density (up to proportionality) of $\log(\sigma)$?

$$\text{Let } G(\sigma) = \log(\sigma) \Rightarrow G^{-1}(\sigma) = e^{\sigma}$$

$$dG^{-1} = e^{\sigma}$$

$$f_{\sigma}(\sigma) \propto \frac{1}{e^{\sigma}}|e^{\sigma}| = 1$$

Change of Variables Example

Recall

$$f_Y(y) = f_Z(G^{-1}(y))|\det(dG^{-1})|$$

Suppose σ has probability density $f(\sigma) \propto 1/\sigma$. What is the probability density (up to proportionality) of $\log(\sigma)$?

$$\text{Let } G(\sigma) = \log(\sigma) \Rightarrow G^{-1}(\sigma) = e^{\sigma}$$

$$dG^{-1} = e^{\sigma}$$

$$f_{\sigma}(\sigma) \propto \frac{1}{e^{\sigma}}|e^{\sigma}| = 1$$

Change of Variables Example

Recall

$$f_Y(y) = f_Z(G^{-1}(y))|\det(dG^{-1})|$$

Suppose σ has probability density $f(\sigma) \propto 1/\sigma$. What is the probability density (up to proportionality) of $\log(\sigma)$?

$$\begin{aligned} \text{Let } G(\sigma) = \log(\sigma) &\Rightarrow G^{-1}(\sigma) = e^{\sigma} \\ dG^{-1} &= e^{\sigma} \\ f_{\sigma}(\sigma) &\propto \frac{1}{e^{\sigma}}|e^{\sigma}| = 1 \end{aligned}$$

Expectation of a Random Variable

1. Discrete
   - Suppose X is a discrete random variable taking on values $x_1, x_2, \ldots$ with probabilities $p_1, p_2, \ldots$. Then the expected value of X and a function of X, $f(X)$ is

$$\sum_{i=1}^{\infty} x_i p_i \quad / \quad \sum_{i=1}^{\infty} f(x_i) p_i$$

2. Continuous
   - Suppose X is a continuous random variable with probability density function $g(X)$. Then the expected value of X and a function of X, $f(X)$ is

$$\int_A x g(x) dx \quad / \quad \int_A f(x) g(x) dx$$

   where A is the support of the random variable.

Expectation Example

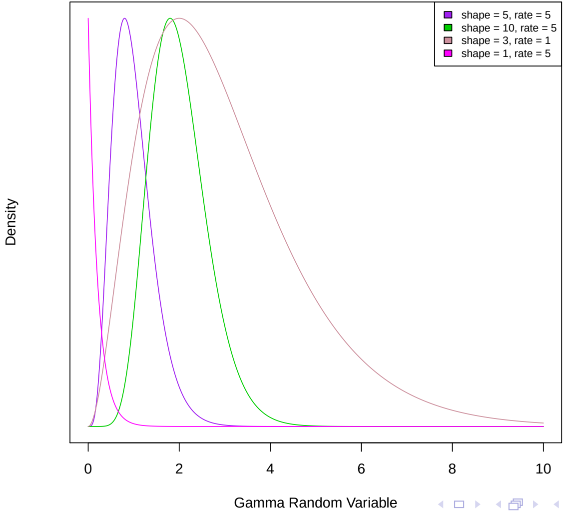
The density function of a Gamma distributed random variable X with shape parameter α and rate parameter β is

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

What is the expected value of X?

Expectation Example

Gamma Densities at Various Shape and Rate Values:



Expectation Example

$$
\begin{aligned}
E[X] &= \int_0^\infty x \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{(\alpha+1)-1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \int_0^\infty \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} x^{(\alpha+1)-1} e^{-\beta x} dx \\
&= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\beta} = \frac{\alpha\Gamma(\alpha)}{\Gamma(\alpha)\beta} = \frac{\alpha}{\beta}
\end{aligned}
$$

Expectation Example

$$
\begin{aligned}
E[X] &= \int_0^\infty x \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{(\alpha+1)-1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \int_0^\infty \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} x^{(\alpha+1)-1} e^{-\beta x} dx \\
&= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\beta} = \frac{\alpha\Gamma(\alpha)}{\Gamma(\alpha)\beta} = \frac{\alpha}{\beta}
\end{aligned}
$$

Expectation Example

$$
\begin{aligned}
E[X] &= \int_0^\infty x \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{(\alpha+1)-1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \int_0^\infty \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} x^{(\alpha+1)-1} e^{-\beta x} dx \\
&= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\beta} = \frac{\alpha\Gamma(\alpha)}{\Gamma(\alpha)\beta} = \frac{\alpha}{\beta}
\end{aligned}
$$

Expectation Example

$$
\begin{aligned}
E[X] &= \int_0^\infty x \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{(\alpha+1)-1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \int_0^\infty \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} x^{(\alpha+1)-1} e^{-\beta x} dx \\
&= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)\beta} = \frac{\alpha\Gamma(\alpha)}{\Gamma(\alpha)\beta} = \frac{\alpha}{\beta}
\end{aligned}
$$

Iterated Expectation and Iterated Variances

Assume that you have two random variables X and Y, such that X is integrable i.e. $E[|X|] < \infty$ and Y not necessarily integrable, but on the same probability space $(\Omega, \mathcal{F}, P)$ then:

$$E[X] = E[E[X|Y]]$$

and

$$V[X] = E[V[X|Y]] + V[E[X|Y]]$$

Iterated Expectation and Variance Example

Assume that you are told that $Y \sim \mathrm{N}(\mu, \sigma^2)$ and $X|Y \sim \mathrm{N}(my, t^2)$. Using **Iterated Expectation** and **Variances** we can find the marginal distribution of X:

$$E[X] = E[E[X|Y]] =$$
$$E[my] = mE[y] = m\mu$$

$$V[X] = E[V[X|Y]] + V[E[X|Y]] =$$
$$E[t^2] + V[my] =$$
$$t^2 + m^2 V[y] =$$
$$t^2 + m^2\sigma^2$$

Hence $X \sim \mathrm{N}(m\mu, t^2 + m^2\sigma^2)$

Iterated Expectation and Variance Example

Assume that you are told that $Y \sim \mathrm{N}(\mu, \sigma^2)$ and $X|Y \sim \mathrm{N}(my, t^2)$. Using **Iterated Expectation** and **Variances** we can find the marginal distribution of X:

$$E[X] = E[E[X|Y]] =$$
$$E[my] = mE[y] = m\mu$$

$$\begin{aligned} V[X] = E[V[X|Y]] + V[E[X|Y]] &= \\ E[t^2] + V[my] &= \\ t^2 + m^2 V[y] &= \\ t^2 + m^2\sigma^2 & \end{aligned}$$

Hence $X \sim \mathrm{N}(m\mu, t^2 + m^2\sigma^2)$

Iterated Expectation and Variance Example

Assume that you are told that $Y \sim \mathrm{N}(\mu, \sigma^2)$ and $X|Y \sim \mathrm{N}(my, t^2)$. Using **Iterated Expectation** and **Variances** we can find the marginal distribution of X:

$$E[X] = E[E[X|Y]] =$$
$$E[my] = mE[y] = m\mu$$

$$\begin{aligned} V[X] = E[V[X|Y]] + V[E[X|Y]] &= \\ E[t^2] + V[my] &= \\ t^2 + m^2 V[y] &= \\ t^2 + m^2\sigma^2 & \end{aligned}$$

Hence $X \sim \mathrm{N}(m\mu, t^2 + m^2\sigma^2)$

Bayes' Rule

Given two events A and B, with marginal probabilities $P(A)$ and $P(B)$, we have that

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Given two random variables θ and X,

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int_\theta f(x|\theta)f(\theta)d\theta} \quad \Big/ \quad f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\sum_\theta f(x|\theta)f(\theta)d\theta}$$

Bayes' Rule Example

A random woman over the age of 50 tests positive for breast cancer during her mammogram. It is known that 1 percent of women over 50 have breast cancer, 90 percent of women who have breast cancer test positive on mammograms, and 8 percent of women who do not have breast cancer falsely test positive. What is the posterior probability that the woman has breast cancer?

Let C^+ and C^- represent the event that the woman has breast cancer and the event that woman does not have breast cancer respectively.

$$P(C^+|+) = \frac{P(+|C^+)P(C^+)}{P(+)} = \frac{.9(.01)}{.9(.01) + .08(.99)}$$
$$\approx 0.102$$

Bayes' Rule Example

A random woman over the age of 50 tests positive for breast cancer during her mammogram. It is known that 1 percent of women over 50 have breast cancer, 90 percent of women who have breast cancer test positive on mammograms, and 8 percent of women who do not have breast cancer falsely test positive. What is the posterior probability that the woman has breast cancer?

Let C^+ and C^- represent the event that the woman has breast cancer and the event that woman does not have breast cancer respectively.

$$P(C^+|+) = \frac{P(+|C^+)P(C^+)}{P(+)} = \frac{.9(.01)}{.9(.01) + .08(.99)}$$
$$\approx 0.102$$

The Central Limit Theorem

Assume that $X_1, \ldots, X_n$ be a random sample of size n that is an IID sequence drawn from distributions with finite expectation (μ) and variance (σ^2). Then:

$$S_n = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$S_n \xrightarrow{d} \mathrm{N}(\mu, \frac{\sigma^2}{n})$$

Central Limit Theorem Example

Assume that you have $X_1, \ldots, X_{1000} \sim \mathrm{Poisson}(\lambda)$, what is the approximate distribution of $S_n = \frac{\sum_{i=1}^{n=100} X_i}{100}$?

Since n is large (usually > 30), and based on the poisson distribution we know that $E[S_{100}] = \lambda$ and $V[S_{100}] = \frac{\lambda}{100}$. Then invoking the Central Limit Theorem:

$$S_{100} \overset{\text{Approx}}{\sim} \mathrm{N}(\lambda, \frac{\lambda}{100})$$

Central Limit Theorem Example

Assume that you have $X_1, \ldots, X_{1000} \sim \mathrm{Poisson}(\lambda)$, what is the approximate distribution of $S_n = \frac{\sum_{i=1}^{n=100} X_i}{100}$?

Since n is large (usually > 30), and based on the poisson distribution we know that $E[S_{100}] = \lambda$ and $V[S_{100}] = \frac{\lambda}{100}$. Then invoking the Central Limit Theorem:

$$S_{100} \overset{\mathrm{Approx}}{\sim} \mathrm{N}(\lambda, \frac{\lambda}{100})$$

Central Limit Theorem Example

Assume that you have $X_1, \ldots, X_{1000} \sim \mathrm{Poisson}(\lambda)$, what is the approximate distribution of $S_n = \frac{\sum_{i=1}^{n=100} X_i}{100}$?

Since n is large (usually > 30), and based on the poisson distribution we know that $E[S_{100}] = \lambda$ and $V[S_{100}] = \frac{\lambda}{100}$. Then invoking the Central Limit Theorem:

$$S_{100} \overset{\text{Approx}}{\sim} \mathrm{N}(\lambda, \frac{\lambda}{100})$$

Matrix Properties

- Any matrix with the same number of rows and columns is called a *square matrix*.
- Let $A = [a_{ij}]$ be a matrix. The transpose of A, written as $A^{'}$ or A^T is the matrix $A^T = [b_{ij}]$, where $b_{ij} = a_{ji}$
- If $A = A^T$, then A is called symmetric. Note that only square matrices can be symmetric.
- If A is a square matrix $[a_{ij}]$ and $a_{ij} = 0$ for $i \neq j$, then A is a *diagonal* matrix.
- Let A be a square $n \times n$ matrix. A is nonsingular if there exists a matrix A^{-1} such that $A^{-1}A = I_n = AA^{-1}$.

Matrix Properties

- Any matrix with the same number of rows and columns is called a *square matrix*.
- Let $A = [a_{ij}]$ be a matrix. The transpose of A, written as $A^{'}$ or A^T is the matrix $A^T = [b_{ij}]$, where $b_{ij} = a_{ji}$
- If $A = A^T$, then A is called symmetric. Note that only square matrices can be symmetric.
- If A is a square matrix $[a_{ij}]$ and $a_{ij} = 0$ for $i \neq j$, then A is a *diagonal* matrix.
- Let A be a square $n \times n$ matrix. A is nonsingular if there exists a matrix A^{-1} such that $A^{-1}A = I_n = AA^{-1}$.

Matrix Properties

- Any matrix with the same number of rows and columns is called a *square matrix*.
- Let $A = [a_{ij}]$ be a matrix. The transpose of A, written as $A^{'}$ or A^{T} is the matrix $A^{T} = [b_{ij}]$, where $b_{ij} = a_{ji}$
- If $A = A^{T}$, then A is called symmetric. Note that only square matrices can be symmetric.
- If A is a square matrix $[a_{ij}]$ and $a_{ij} = 0$ for $i \neq j$, then A is a *diagonal* matrix.
- Let A be a square $n \times n$ matrix. A is nonsingular if there exists a matrix A^{-1} such that $A^{-1}A = I_n = AA^{-1}$.

Matrix Properties

- Any matrix with the same number of rows and columns is called a *square matrix*.
- Let $A = [a_{ij}]$ be a matrix. The transpose of A, written as A' or A^T is the matrix $A^T = [b_{ij}]$, where $b_{ij} = a_{ji}$
- If $A = A^T$, then A is called symmetric. Note that only square matrices can be symmetric.
- If A is a square matrix $[a_{ij}]$ and $a_{ij} = 0$ for $i \neq j$, then A is a *diagonal* matrix.
- Let A be a square $n \times n$ matrix. A is nonsingular if there exists a matrix A^{-1} such that $A^{-1}A = I_n = AA^{-1}$.

Matrix Properties

- A matrix P is nonsingular if and only if the columns of A are linearly independent (i.e. no column is a linear combination of the other columns)
- A square matrix P is orthogonal if $P^T = P^{-1}$
- $C(X)$ denotes the column space of a matrix X which is the space spanned by the linearly independent columns in the matrix.
- $N(X)$ denotes the null space of a matrix which are all non-zero column vectors A that satisfy the equation $XA = \vec{0}$
- $r(X)$ denotes the rank of a matrix i.e. the number of linear independent columns in a matrix.

Matrix Properties

- A matrix P is nonsingular if and only if the columns of A are linearly independent (i.e. no column is a linear combination of the other columns)
- A square matrix P is orthogonal if $P^T = P^{-1}$
- $C(X)$ denotes the column space of a matrix X which is the space spanned by the linearly independent columns in the matrix.
- $N(X)$ denotes the null space of a matrix which are all non-zero column vectors A that satisfy the equation $XA = \vec{0}$
- $r(X)$ denotes the rank of a matrix i.e. the number of linear independent columns in a matrix.

Matrix Properties

- A matrix P is nonsingular if and only if the columns of A are linearly independent (i.e. no column is a linear combination of the other columns)
- A square matrix P is orthogonal if $P^T = P^{-1}$
- $C(X)$ denotes the column space of a matrix X which is the space spanned by the linearly independent columns in the matrix.
- $N(X)$ denotes the null space of a matrix which are all non-zero column vectors A that satisfy the equation $XA = \vec{0}$
- $r(X)$ denotes the rank of a matrix i.e. the number of linear independent columns in a matrix.

Matrix Properties

- A matrix P is nonsingular if and only if the columns of A are linearly independent (i.e. no column is a linear combination of the other columns)
- A square matrix P is orthogonal if $P^T = P^{-1}$
- $C(X)$ denotes the column space of a matrix X which is the space spanned by the linearly independent columns in the matrix.
- $N(X)$ denotes the null space of a matrix which are all non-zero column vectors A that satisfy the equation $XA = \vec{0}$
- $r(X)$ denotes the rank of a matrix i.e. the number of linear independent columns in a matrix.

Matrix Properties Continued

- The Basis for a space N is a collection of linearly independent vectors that span N.
- A symmetric matrix A is positive definite if for any non-zero vector $v \in \mathbb{R}^n$, $v^T Av > 0$
- A matrix P is idempotent if $P \cdot P = P$

Useful Resources

- Plane Answers to Complex Questions - *Christensen* 2011
- Statistical Inference - *Casella & Berger* 2002
- Wikipedia, Stack Exchange, Google, Bing(jk...), Yandex (really jk...), Baidu (do you even have to ask?)