ESE 524: Probability Review and Useful Tools

January 17, 2019

Probability Overview

- Assume X is a random variable with domain D (e.g. $\mathbb{R}$).
- The Expected Value is $\mathbb{E}[X] = \int_D xp(x)dx$ for continuous variables, and $\sum_D xp(x)$ for discrete variables.
- The variance of X is $\mathrm{var}(x) = E[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
- Marginal Distributions: Given random variables X and Y with domains D_X and D_Y, a joint probability distribution $p(x, y)$ the marginal distribution $p(x) = \int_{D_Y} p(x, y)dy$.
- Conditional Distributions: The conditional probability distribution of X given Y is $p(x|y) = \frac{p(x,y)}{p(x)}$.
- Conditional Expectation: expected value of the conditional distribution, $\mathbb{E}_{X|Y}[X|Y] = \int_{D_X} xp(x|y)dx$.

Managing Expectations

- Assume X and Y are continuous random variables.
- The expected value operation is linear, i.e.:

$$\mathbb{E}[5X + 6Y] = 5\mathbb{E}[X] + 6\mathbb{E}[Y]$$

- Let's prove some of the expected value formulas from L1 on the board:
  - $\mathbb{E}_Y[\mathbb{E}_{X|Y}[X|Y]] = \mathbb{E}_X[X]$
  - $\mathbb{E}_{X|Y}[g(X)h(Y)|Y = y] = h(y)\mathbb{E}_{X|Y}[g(X)|Y = y]$
  - $\mathrm{var}_X(X) = \mathbb{E}_Y[\mathrm{var}_{X|Y}(X|Y)] + \mathrm{var}_Y(\mathbb{E}_{X|Y}[X|Y])$
  - $\mathrm{cov}(\mathbf{X}) = \mathbb{E}[\mathrm{cov}(\mathbf{X}|Y)] + \mathrm{cov}(\mathbb{E}[\mathbf{X}|\mathbf{Y}])$
  - $\mathbb{E}_{X,Y}[g(X)h(Y)] = \mathbb{E}_Y[h(Y)\mathbb{E}_{X|Y}[g(X)|Y]]$
- Why is this important?
- When can you change the order of integration?

Transformation of Random Variables Example

- Given:
  - Random variable X with pdf $p_x(x) = x\exp(\frac{-x^2}{2})$
  - Standard Normal Random variable Y with pdf $p_y(y) = \frac{1}{\sqrt{2\pi}}\exp(\frac{-y^2}{2})$
  - constant c

  Find the joint distribution of $U = g_1(x, y) = \sqrt{X^2 + Y^2}$ and $V = g_2(x, y) = \frac{cY}{X}$.
- First solve for the inverse transformation (see board for details):

$$x = h_1(u, v) = \frac{u}{\sqrt{1 + \frac{v^2}{c^2}}}$$

$$y = h_2(u, v) = \frac{uv}{c\sqrt{1 + \frac{v^2}{c^2}}}$$

- Then find the Jacobian matrix:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{1+v^2/c^2}} & \frac{-uv}{c^2(1+v^2/c^2)^{3/2}} \\ \frac{v}{c\sqrt{1+v^2/c^2}} & \frac{u}{c(1+v^2/c^2)^{3/2}} \end{bmatrix}$$

Transformation of Random Variables Example Continued

- The determinant of $\mathbf{J}$ is $\det(\mathbf{J}) = \frac{cu}{c^2+v^2}$
- Following the transformation of variables formula:

$$p_{u,v}(u,v) = p_{x,y}(h_1(u,v), h_2(u,v))|\det(\mathbf{J})| = \frac{u}{\sqrt{1+v^2/c^2}} \frac{e^{-u^2/2}}{\sqrt{2\pi}} \frac{cu}{c^2+v^2}$$

- After some algebra, we can separate $p_{u,v}$ into:

$$p_{u,v} = \sqrt{\frac{2}{\pi}} u^2 e^{-u^2/2} . \frac{1}{2c} \left(1 + \frac{v^2}{c^2}\right)^{-3/2}$$

which is the product of a Maxwell distribution and student's-t distribution.

Common Probability Distributions

Name	Probability Distribution	Mean	Variance	Section in Book
Discrete				
Uniform	$\frac{1}{n}, a \le b$	$\frac{(b+a)}{2}$	$\frac{(b-a+1)^2-1}{12}$	3-5
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$	np	$np(1-p)$	3-6
	$x = 0, 1, \ldots, n, 0 \le p \le 1$			
Geometric	$(1-p)^{x-1} p$ $x = 1, 2, \ldots, 0 \le p \le 1$	$1/p$	$(1-p)/p^2$	3-7
Negative binomial	$\binom{x-1}{r-1}(1-p)^{x-r} p^r$ $x = r, r+1, r+2, \ldots, 0 \le p \le 1$	r/p	$r(1-p)/p^2$	3-7
Hypergeometric	$\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$ $x = \max(0, n-N+K), 1, \ldots$ $\min(K, n), K \le N, n \le N$	np where $p = \frac{K}{N}$	$np(1-p)\left(\frac{N-n}{N-1}\right)$	3-8
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots, 0 < \lambda$	λ	λ	3-9
Continuous				
Uniform	$\frac{1}{b-a}, a \le x \le b$	$\frac{(b+a)}{2}$	$\frac{(b-a)^2}{12}$	4-5
Normal	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $-\infty < x < \infty, -\infty < \mu < \infty, 0 < \sigma$	μ	σ^2	4-6
Exponential	$\lambda e^{-\lambda x}, 0 \le x, 0 < \lambda$	$1/\lambda$	$1/\lambda^2$	4-8
Erlang	$\frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!}, 0 < x, r = 1, 2, \ldots$	r/λ	r/λ^2	4-9.1
Gamma	$\frac{\lambda x^{r-1} e^{-\lambda x}}{\Gamma(r)}, 0 < x, 0 < r, 0 < \lambda$	r/λ	r/λ^2	4-9.2
Weibull	$\frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} e^{-(x/\delta)^\beta}$ $0 < x, 0 < \beta, 0 < \delta$	$\delta\Gamma\left(1+\frac{1}{\beta}\right)$	$\delta^2\Gamma\left(1+\frac{2}{\beta}\right) - \delta^2\left[\Gamma\left(1+\frac{1}{\beta}\right)\right]^2$	4-10
Lognormal	$\frac{1}{x\omega\sqrt{2\pi}} \exp\left(\frac{-[\ln(x)-\theta]^2}{2\omega^2}\right)$	$e^{\theta+\omega^2/2}$	$e^{2\theta+\omega^2}(e^{\omega^2}-1)$	4-11
Beta	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$ $0 \le x \le 1, 0 < \alpha, 0 < \beta$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	4-12

Figure 1: Table of common probability distributions for reference. Retrieved from *Applied Statistics and Probability for Engineers*, by Runger and Montgomery.

Getting a Probability Distribution from a Model

- The general method for solving problems in this class will proceed as follows
  - Given a set of known samples, $\mathbf{x} = [x[0], x[1], ..., x[n-1]]$, and unknown parameter (or vector of parameters) θ, assume a model $x[n] = f(\theta, n)$.

Getting a Probability Distribution from a Model

- The general method for solving problems in this class will proceed as follows
  - Given a set of known samples, $\mathbf{x} = [x[0], x[1], ..., x[n-1]]$, and unknown parameter (or vector of parameters) θ, assume a model $x[n] = f(\theta, n)$.
  - Use the model and the transformation of variables formula to find the likelihood function, $p(x|\theta)$.

Getting a Probability Distribution from a Model

- The general method for solving problems in this class will proceed as follows
  - Given a set of known samples, $\mathbf{x} = [x[0], x[1], ..., x[n-1]]$, and unknown parameter (or vector of parameters) θ, assume a model $x[n] = f(\theta, n)$.
  - Use the model and the transformation of variables formula to find the likelihood function, $p(x|\theta)$.
  - Use the likelihood to construct an estimator $\hat{\theta}(\mathbf{x})$. Note this is another transformation of random variables, so it has a probability distribution.

Getting a Probability Distribution from a Model

- The general method for solving problems in this class will proceed as follows
  - Given a set of known samples, $\mathbf{x} = [x[0], x[1], ..., x[n-1]]$, and unknown parameter (or vector of parameters) θ, assume a model $x[n] = f(\theta, n)$.
  - Use the model and the transformation of variables formula to find the likelihood function, $p(x|\theta)$.
  - Use the likelihood to construct an estimator $\hat{\theta}(\mathbf{x})$. Note this is another transformation of random variables, so it has a probability distribution.
  - Examine the performance of an estimator, usually by computing an expected value such as the Mean-Squared-Error.

Example

- Model: Estimating a DC Level with Additive Noise $x[n] = \theta + w[n]$ where each $w[n]$ is a sample from white noise, and is normally distributed with mean 0 and known variance σ^2.

Example

- Model: Estimating a DC Level with Additive Noise $x[n] = \theta + w[n]$ where each $w[n]$ is a sample from white noise, and is normally distributed with mean 0 and known variance σ^2.
- Likelihood Function: Assuming the samples $x[n]$ are independent, we can make the likelihood function as a multivariate gaussian.

Example

- Model: Estimating a DC Level with Additive Noise $x[n] = \theta + w[n]$ where each $w[n]$ is a sample from white noise, and is normally distributed with mean 0 and known variance σ^2.
- Likelihood Function: Assuming the samples $x[n]$ are independent, we can make the likelihood function as a multivariate gaussian.
- Estimator: Intuitively, the average, $\hat{\theta} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$, is the best estimator here.

Example

- Model: Estimating a DC Level with Additive Noise $x[n] = \theta + w[n]$ where each $w[n]$ is a sample from white noise, and is normally distributed with mean 0 and known variance σ^2.
- Likelihood Function: Assuming the samples $x[n]$ are independent, we can make the likelihood function as a multivariate gaussian.
- Estimator: Intuitively, the average, $\hat{\theta} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$, is the best estimator here.
- Performance: We saw in example 1 from Lecture 1 that the MSE of this estimator is σ^2/N.