Probability and Statistics for Final Year Engineering Students

By Yoni Nazarathy, Last Updated: May 22, 2011.

Exercises and Tutorial 1: Introduction and Basic Terms

Basic "Statistics":

A "statistic" is a quantity calculated from a random sample, $X_1, X_2, \ldots, X_n$. Here are important examples:

- The minimum or maximum, $Min(X_1, X_2, \ldots, X_n)$ or $Max(X_1, X_2, \ldots, X_n)$.
- The sample mean, $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$.
- The sample variance, $S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$
- The proportion of values satisfying some property A: $\hat{p} = \frac{\sum_{i=1}^{n} 1\{X_i \in A\}}{n}$.

1. Assume$(X_1, X_2, \ldots, X_6) = (-10.3, 7.2, 4.2, 1.2, 0,\ -12)$, calculate the above statistics. For the proportion statistic let $A = \{x : x < 0\}$.
2. Assume that the $X_i = 3i + 7$ for $i = 1, \ldots, n$ (this is obviously not a "random sample" – yet assume this for the purpose of the exercise). Calculate the above statistics, for the proportion statistic let A be the set of even integers.
3. Find a formula for the sample mean of n+1 observations given X_{n+1} and the sample mean of n observations. When is this useful?
4. Find a formula for the sample variance in terms of the sample mean and $\sum_{i=1}^{n} {X_i}^2$ (i.e. the formula should only use these two values and n, but not the individual observations).
5. Assume that the data is binary taking values which are either 0 or 1. The sample proportion is $\hat{p} = \frac{\sum_{i=1}^{n} I_i}{n}$. Write a formula for the sample variance in terms of $\hat{p}$.
6. What is $\sum_{i=1}^{n} (X_i - \bar{X})$?

Probability and random variables:

The cumulative distribution function (CDF) of a random variable X (discrete or continuous) is:

$$F(x) = P(X \le x).$$

The PDF (density) of a continuous random variable is: $f(x) = \frac{d}{dx}F(x)$. We have that

$$P(a \le X \le b) = \int_a^b f(x)dx = F(b) - F(a).$$

The PDF (mass – sometimes called PMF) of a discrete random variable is

$$p(x) = P(X = x) = F(x) - F(x^-).$$

The mean (expectation) of a random variable is:

$$E[X] = \sum_x x\, p(x) \qquad\qquad E[X] = \int x f(x)dx,$$

And the variance is:

$$Var(X) = \sum_x (x - E[X])^2\, p(x) \qquad\qquad Var(X) = \int (x - E[X])^2 f(x)dx.$$

If $Y = g(X)$, one way to calculate the mean and variance of Y is to find its distribution (we don't fully do this in this subject). Alternatively: $E[Y] = \sum_x g(x)\, p(x)\ \ and\ Var(Y) = \sum_x (g(x) - E[Y])^2\, p(x)$. (With similar formulas for the continuous case).

7. Write down and draw the CDF and PDF of a uniform random variable taking values in the range [-2,8].
8. Calculate the mean and variance of the above random variable.
9. Show that $Var(X) = E[X^2] - (E[X])^2$.
10. A discrete uniform random variable gets values $\{1, \dots, n\}$, each with probability $1/n$. Calculate the mean and the variance of this random variable.
11. Consider a continuous random variable with $f(x) = Cx^2$ for $-1 \le x \le 1$.
    a. What is the value of C?
    b. What is the mean?
    c. What is the variance?
12. A "degenerate" random variable gets the value c with probability 1. What is the CDF of this random variable? What is the mean? What is the variance?
13. An exponential random variable X has PDF, $f(x) = \frac{1}{\mu} e^{-\frac{1}{\mu}x}$ for $x \ge 0$, where $\mu > 0$. Assume $\mu = 5.2$ and calculate $P(X > x)$.
14. Show that if X is a random variable, then for the random variable Y=aX+b:

$$E[Y] = aE[X] + b \qquad\qquad Var(Y) = a^2 Var(X).$$

The distribution of a statistic:
A statistic is a random variables which is a function of the random sample.

Here are basic properties of the mean and variance which we need:

$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots E[X_n] \quad ,$$

for any sequence of random variables. When the sequence is **independent** then,

$$Var(X_1 + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n).$$

(Note: In section 4 we will learn about covariance which is used to correct the variance of the sum of dependent random variables).

15. Show that $E[\bar{X}] = E[X]$. Does the result require the random variables in the sample to be independent?
16. Show that $Var(\bar{X}) = \frac{Var(X)}{n}$. Does the result require the random variables in the sample to be independent?
17. Assume $X_1, \ldots, X_n$ are independent uniform random variables on the range [10,20]. What is the mean and variance of the sample mean?
18. Assume $I_1, \ldots, I_n$ are independent Bernoulli random variables with probability p (each gets 1 with probability p and 0 with probability 1-p) , what is the mean and variance of the sample mean (which is also the sample proportion)?
19. Use simulation to obtain the CDF of the maximum of 3 independent uniform (0,1) random variables. Compare with this calculation:

$$P(Max(X_1, \ldots, X_n) \leq x) = P(X_1 \leq x, \ldots, X_n \leq x) = \mathrm{F(x)}^n = \begin{cases} 0 & x < 0 \\ x^n & 0 \leq x \leq 1. \\ 1 & 1 < x \end{cases}$$

20. Try to use similar reasoning to that of the above calculation to obtain the distribution of $Min(X_1, \ldots, X_n)$.
21. What is the mean value (expectation) of $Min(X_1, \ldots, X_n)$ in the case of independent uniform (0,1) random variables?
22. Use the above result to calculate the mean and variance of the sample maximum in the case of uniform (0,1) for arbitrary n.

Selected Solutions

1) Sample mean: -1.62 sample variance: 61.1, $\hat{p} = 1/3$.
2) Sample mean: $\frac{17n+3n^2}{2n}$ Sample variance: $\frac{3(n^3-n)}{4(n-1)}$ (a bit tedious). Sample proportion: if n is even then ½ if n is odd then $\frac{n+1}{2n}$.
3) $\bar{X}_{n+1} = \frac{n}{n+1}\bar{X}_n + \frac{1}{n+1}X_{n+1}$. This is useful for an "on-line running average".
4) See Lecture 3:

$$S^2 = \frac{\sum_{i=1}^{n} X_i^{\,2} - n\bar{X}^2}{n-1}.$$

5) Observe that $n\hat{p} = \sum_{i=1}^{n} I_i = \sum_{i=1}^{n} I_i^{\,2}$. So

$$S^2 = \frac{n\hat{p}-n\hat{p}^2}{n-1} = \frac{n}{n-1}\hat{p}(1-\hat{p}).$$

6) 0.
7) CDF: $F(x) = \begin{cases} 0 & x < -2 \\ \frac{1}{10}x + \frac{1}{5} & -2 \le x \le 8 \\ 1 & 8 < x \end{cases}$ PDF: $f(x) = \begin{cases} 0 & x < -2 \\ \frac{1}{10} & -2 \le x \le 8. \\ 0 & 8 < x \end{cases}$
8) Mean, E[X]=$\int_{-\infty}^{\infty} xf(x)dx = \int_{-2}^{8} x\frac{1}{10}dx = 3$

   Variance: Var(X)= $\int_{-\infty}^{\infty}(x-3)^2 f(x)dx = \int_{-2}^{8}(x^2-6x+9)\frac{1}{10}dx = \frac{25}{3}$
9) This is a very useful result: (denote $EX = \mu$). Below we show the result is for the discrete case, for the continuous case replace the sums by integrals.

$$Var(X) = \sum (x-\mu)^2 p(x) = \sum (x^2 - 2x\mu + \mu^2)p(x) = \sum x^2 p(x) - 2\mu^2 + \mu^2$$

10) Mean: $\frac{1+n}{2}$. Variance: $\frac{n^2-1}{12}$
11) a) C=3/2. b) 0 c) $\int_{-1}^{1} \frac{3}{2} x^4 dx = \frac{3}{5}$
12) CDF is the step function at the point c (gets 0 for values less than c and 1 for values greater or equal to c). The mean is c. The variance is 0.
13) $e^{-0.192308x}$
14) $E[aX+b] = \int (ax+b)f(x)dx = a\int xf(x)dx + b\int f(x)dx = aE[X] + b$

$$Var(aX+b) = \int (aX + b - E[aX+b])^2 f(x)dx = \int (aX + b - aE[X] - b)^2 f(x)dx = a^2 Var(X)$$

The calculation for discrete random variables follows the exact same lines.

15) $E[\bar{X}] = E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{1}{n}E[\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{nE[X]}{n} = E[X]$. This is true whenever the random variables have the same mean yet does not require them to be independent.
16) $Var(\bar{X}) = Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2}Var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{nVar(X)}{n^2} = \frac{Var(X)}{n}$

   This is true whenever the variance of a sum is the sum of the variances, as is the case when the random variables are independent.
17) $E[\bar{X}] = 15, Var(\bar{X}) = \frac{25}{3n}$.

18) QQQQ

19) QQQQ

20) $F(x) = P(Min(X_1, \ldots, X_n) \leq x) = 1 - P(Min(X_1, \ldots, X_n) > x)$

$$= 1 - P(X_1 > x, \ldots, X_n > x) = 1 - P(X_1 > x) \ldots P(X_n > x) = 1 - (1 - F(x))^n$$

21) Using the above, $F(x) = 1 - (1 - x)^n$ for x in [0,1]. So the density is:

$$f(x) = n(1 - x)^{n-1}$$

So the mean is $\int_0^1 xn(1 - x)^{n-1}dx = \frac{1}{1+n}$