Stat 135 Probability review sheet   Summer 2018

1 **Formulas**

a **Baye's rule** $Pr(A|B) = \frac{Pr(A\cap B)}{Pr(B)}$ or $Pr(A \cap B) = Pr(A|B)Pr(B)$ (multiplication rule)

b **Independence of two events** $Pr(A|B) = Pr(A)$ or $Pr(A \cap B) = Pr(A)Pr(B)$.

c **Independence of two RVs (discrete case)** $Pr(X = x|Y = y) = Pr(X = x)$ or $Pr(X = x, Y = y) = Pr(X = x)Pr(Y = y)$

d **Expectation of a function (continuous case)**: Let X be continuous RV with density f and let $g(X)$ be a function of X. Then $E(g(X)) = \int_{x\in X} g(x)f(x)dx$. Important special case is

$$E(X) = \int_{x\in X} xf(x)dx$$

e **Variance** $Var(X) = E[(X - E(X))^2]$ or $Var(X) = E[X^2] - (E[X])^2$.

f **Standard Deviation** $SD(X) = \sqrt{Var(X)}$

g **Markov's inequality** $Pr(X \geq c) \leq \frac{E(X)}{c}$

h **Chebyshev's inequality** $Pr(|X - E(X)| \geq zSD[X]) \leq \frac{1}{z^2}$

i **Expectation of a product of RV (discrete case)** $E[XY] = \sum_{x\in X}\sum_{y\in Y} xyPr(X = x, Y = y)$

j **Covariance of two RV** $Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$

k **Correlation of two RV** $Corr(X, Y) = \frac{Cov(X,Y)}{SD[X]SD[Y]}$ Note that $-1 \leq Corr(X, Y) \leq 1$ and is correlation is unitless.

l **Variance of a sum of RV** $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$. In the special case where $X_1, \dots X_n$ are identically distributed (not necessarily independent),

$$Var(X_1 + \cdots + X_n) = nVar(X_1) + n(n-1)Cov(X_1, X_2).$$

We have used this formulas for indicators!

m **Generation of pseudo random numbers having a specified distribution** $X = F_X^{-1}(U)$ where F_X is the cummulative distribution function (cdf) of X, and U is the standard uniform RV.

n **Change of Variables formula** If $Y = g(X)$ then $f_Y(y) = \frac{f_X(x)}{|g'(x)|}$ where $x = g^{-1}(y)$. Remember that if the function $g(X)$ is 2-1 (two to one) that the change of variable formula has two terms, one for each value of y for a given x.

o **conditional density** $f(y|x) = \frac{f(x,y)}{f_X(x)}$ where $f(x, y)$ is the joint density for two RVs (X,Y), and $f_X(x) = \int_{y\in Y} f(x, y)dy$ is the marginal denisity.

p **conditional expectation** $E(Y|X) = \int_{y\in Y} yf(y|x)dy$

q **Convolution formula for density of Z=X+Y (continous case)** $f_Z(z) = \int_{x=-\infty}^{x=\infty} f(x, x - z)dx$. Important special case for $X > 0$ and $Y > 0$ independent is

$$f_Z(z) = \int_{x=0}^{x=z} f_X(x)f_Y(z - x)dx$$

2 Distributions

a Binomial (discrete) Generalizes Bernouilli trial (a.k.a. indicator) which is $Bin(1, p)$. Approximates normal distribution, $N(np, \sqrt{np(1-p)})$, for large n.

examples:

$X \sim Bin(n, p)$: X= number of heads in n coin tosses of a coin which has prob p of being heads. X takes values between 0 and n.

$X \sim Bin(5, 1/13)$: X= number of aces you get when you draw with replacement 5 cards from a deck

$X = I_1 + \cdots + I_n$ where I_j is an indicator for the jth toss being heads. Note: indicators independent

$$Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$E(X) = np$

$SD(X) = \sqrt{np(1-p)}$

b Poisson (discete) Bin(n,p) approximates Pois(np) for large n and small p.

$X \sim Pois(\lambda)$: X= number of cars passing through a toll booth in an hour where λ is the average rate of cars arriving at the toll booth.

$$Pr(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

$E[X] = \lambda$

$SD[X] = \sqrt{\lambda}$

If $Y \sim Pois(\lambda)$ and $Z \sim Pois(\theta)$ are independent then $Y + Z \sim Pois(\lambda + \theta)$

c Hypergeometric (discrete) This is essentially the binomial except that you draw without replacement so the indicators aren't independent

$X \sim Hypergeom(N, G, n)$: X= number of good cards in a deck of N cards having G good cards when you draw without replacement a sample of size n. So for example X= number of aces you get in a deck of 52 cards when you draw without replacement 5 cards. X takes values between 0 and n.

$X = I_1 + \cdots + I_n$ where I_j is an indicator for the jth card being good. Note: indicators not independent

$$Pr(X = g) = \frac{\binom{G}{g}\binom{N-G}{n-g}}{\binom{N}{n}}$$

$E[X] = \frac{nG}{N}$ compare with binomial

$SD[X] = \sqrt{n \frac{G}{N} \frac{N-G}{N} (\frac{N-n}{N-1})}$ compare with binomial

d Uniform on $\{x_1, \ldots, x_N\}$ (discrete) Think of picking a number out of hat. Compare with uniform distribution on an interval (a,b).

$X \sim Uniform(x_1, \ldots, x_N)$: X= a randomly choosen number from the list of numbers $\{x_1, \ldots, x_N\}$. Note that the numbers can repeat. For example $x_1 = 1, x_2 = 4, x_3 = 1$.

$$Pr(X = x_i) = \frac{1}{N}$$

$E[X] = \frac{\sum_{i=1}^{i=N} x_i}{N}$

$SD[X] = \sqrt{\frac{\sum_{i=1}^{i=N}(x_i - E[x])^2}{N}}$

Special case: Discrete uniform RV $X \sim Unif\{1, 2, \ldots, N\}$ then

$E[X] = \frac{N+1}{2}$

$SD[X] = \sqrt{\frac{N^2-1}{12}}$

e Geometric (discrete) Compare with Exponential distribution

$X \sim Geom(p)$: X= number of independent trials until your first success (with prob p for success). X takes values $1, 2, \ldots$

$$Pr(X = k) = (1 - p)^{k-1}p$$

$E[X] = \frac{1}{p}$

$SD[X] = \frac{\sqrt{1-p}}{p}$

f Negative Binomial (discrete) Generalizes Geometric distribution (r=1). Compare with Gamma distribution

$X \sim Negbin(r, p)$: X= number of independent trials until your *rth* success (with prob p for success).

$$Pr(X = k) = \binom{k-1}{r}(1 - p)^k p^r$$

Out of k-1 trials (we exclude the last one which is a success) choose r-1

$X = T_1 + \cdots + T_r$ where $T_1, T_2, \ldots, T_r$ are i.i.d Geom(p).

$E[X] = r\frac{1}{p}$ compare with geometric

$SD[X] = \frac{\sqrt{r(1-p)}}{p}$ compare with geometric

g Uniform on interval (a,b) (continuous)

$X \sim U(a, b)$: X= a randomly chosen number between a and b.

$$f(x) = \frac{1}{b - a}$$

Important special case is the standard uniform $U(0, 1)$ which has density f(x)=1.

$E[X] = \frac{a+b}{2}$

$SD[X] = \frac{b-a}{\sqrt{12}}$

h Beta (continous) Takes values between 0 and 1. Generalizes the standard uniform distribution which is Beta(1,1).

$X \sim Beta(k+1, n-k+1)$: X= the distribution of a coin flipping heads after you flip a coin n times and get k heads. (this is called a posterior distribution)

$X \sim Beta(k, n-k+1)$: X= the unknown distribution of the kth order statistics, $U_{(k)}$ of n i.i.d. standard uniforms $U_1(0,1), \ldots, U_n(0,1)$ (this is how we introduced this distribution). Special cases: $Beta(n,1)$ is the distribution of $U_{(n)} = max(U_1, \ldots, U_n)$ and $Beta(1,n)$ is the distribution of $U_{(1)} = min(U_1, \ldots, U_n)$.

$X \sim Beta(r,s)$ for $r > 0$, $s > 0$ is a popular model for random variables that take values between 0 and 1.

$$f(x) = \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} x^{r-1}(1-x)^{s-1}, \text{ where } \Gamma(r) = \int_0^\infty x^{r-1}e^{-x}dx$$

Variable part of denisty is $x^{r-1}(1-x)^{s-1}$
$E[X] = \frac{r}{r+s}$
$SD[X] = \sqrt{\frac{rs}{(r+s)^2(r+s+1)}}$

i Normal (continous) Takes values between minus and positive infinity. Famous bell curve. Most important RV in statistics because of the Central Limit theorem. Because of the Central Limit theorem the normal distribution is an approximation of of the binomial distribution for large n. Important special case is the standard normal $N(0,1)$ which has mean 0 and standard deviation 1.

$X \sim N(np, np(1-p))$: X= approximately the number of heads you get when you toss a coin n times with a probability of p for heads.

$X \sim N(\mu, \sigma^2)$: X= the x values of a bell curve with mean μ and standard deviation *sigma*. The bell curve is special in that 67% of the data lies within one standard deviation of the mean, 97% lies within two SD of the mean and 99.8% lie within three SD of the mean.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

The variable part is $e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$
Important special case is the standard normal, $N(0,1)$ which has denisty:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$E[X] = \mu$
$SD[X] = \sigma$

j Exponential (continous) Takes values between 0 and infinity. Think of waiting time for something to happen. Continous analogue of the discrete geometric distribution.

$X \sim Exp(\lambda)$: X= waiting time for the arrival of a car at a toll booth where the average rate of cars arriving at the toll booth is λ.

$$f(x) = \lambda e^{-\lambda x}$$

Variable part is $e^{-\lambda x}$
The survival function, $P(X > x) = e^{-\lambda x}$.
$E[X] = \frac{1}{\lambda}$
$SD[X] = \frac{1}{\lambda}$

k χ^2 (pronounced chi-square) with n degrees of freedom (continous). Special case of Gamma RV ($Gamma(\frac{n}{2}, \frac{1}{2})$

$X \sim \chi^2(n)$: The sum of the square of n i.i.d standard normal random variables.

$X = Z_1^2 + \ldots Z_n^2$ where $Z_1, \ldots, Z_n$ are independent $N(0, 1)$ RVs.

Arises in the Goodness of fit test (when you want to test whether your data, for example number of hits in baseball, has a particular discrete distribution, such as the binomial distribution), and Pearson Chi-square test for independence in contingency tables (which tests whether two categorical variables,for example level of education and hourly wage, are independent).

$$f(x) = \frac{(\frac{1}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x}$$

where $\Gamma(\frac{n}{2})$ is given by the recursive relation $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ and $\Gamma(r + 1) = r\Gamma(r)$.
The variable part of the density is $x^{\frac{n}{2}-1} e^{-\frac{1}{2}x}$.
$E[X] = n$
$SD[X] = \sqrt{2n}$

l Gamma (continous) Takes values between 0 and infinity. Think of waiting time for the *rth* occurence of something to happen. Generalizes exponential distribution (which is $Gamma(1, \lambda)$ and $\chi^2(n)$ distribution (which is $Gamma(\frac{n}{2}, \frac{1}{2})$)

$X \sim Gamma(r, \lambda)$: X= waiting time for the *rth* gram of rice to come down a conveyer belt where the average rate of rice moving down the conveyer belt is λ. Here $r, \lambda > 0$.

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, \text{ where } \Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx$$

Variable part of density is $x^{r-1} e^{-\lambda x}$
Important property of Gamma distribution is $X \sim Gamma(r, \lambda)$ and $Y \sim Gamma(s, \lambda)$ are independent then $X + Y \sim Gamma(r + s, \lambda)$.
$E[X] = \frac{r}{\lambda}$
$SD[X] = \frac{\sqrt{r}}{\lambda}$