Some required calculus skills for STAT 542 & STAT 543

Farzad Sabzikar

Department of Statistics

Iowa State University

sabzikar@iastate.edu

August 5, 2021

Outline

- Topics in STAT 542
- Topics in STAT 543
- Required calculus for STAT 542 & STAT 543

Topics in STAT 542

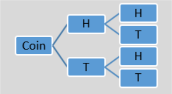
- Probability Theory
- Transformation and Expectations
- Common Families of Distributions
- Multiple Random Variables
- Properties of Random Sample

Topics in STAT 543

- Principles of Data Reduction
- Point Estimation
- Hypothesis Testing
- Interval estimation
- Asymptotic Evaluations

Sample Space

- The set of all possible outcomes of an experiment is called the sample space.



$$S = \left\{ \begin{matrix} (H.H.H).(T.H.H).(H.T.H).(H.H.T) \\ (H.T.T).(T.H.T).(T.T.H).(T.T.T) \end{matrix} \right\}$$

$$S = \left\{ \begin{matrix} (1,1).(1,2).(1,3).(1,4).(1,5).(1,6) \\ (2,1).(2,2).(2,3).(2,4).(2,5).(2,6) \\ (3,1).(3,2).(3,3).(3,4).(3,5).(3,6) \\ (4,1).(4,2).(4,3).(4,4).(4,5).(4,6) \\ (5,1).(5,2).(5,3).(5,4).(5,5).(5,6) \\ (6,1).(6,2).(6,3).(6,4).(6,5).(6.6) \end{matrix} \right\}$$

Random Variable

- A random variable is a real-valued function defined over the elements of a sample space.

Roll a pair of dice and define $\mathbf{X}$ as a total sum of the points.

$$\mathbf{S} = \begin{Bmatrix} (1,1).(1,2).(1,3).(1,4).(1,5).(1,6) \\ (2,1).(2,2).(2,3).(2,4).(2,5).(2,6) \\ (3,1).(3,2).(3,3).(3,4).(3,5).(3,6) \\ (4,1).(4,2).(4,3).(4,4).(4,5).(4,6) \\ (5,1).(5,2).(5,3).(5,4).(5,5).(5,6) \\ (6,1).(6,2).(6,3).(6,4).(6,5).(6.6) \end{Bmatrix}$$

$$X = \begin{cases} 2 & if\ \omega \in \{(1{,}1)\} \\ 3 & if\ \omega \in \{(1{,}2)(2{,}1)\} \\ 4 & if\ \omega \in \{(1{,}3)(2{,}2)(3{,}1)\} \\ 5 & if\ \omega \in \{(1{,}4)(2{,}3)(3{,}2)(4{,}1)\} \\ 6 & if\ \omega \in \{(1{,}5)(2{,}4)(3{,}3)(4{,}2)(5{,}1)\} \\ 7 & if\ \omega \in \{(1{,}6)(2{,}5)(3{,}4)(4{,}3)(5{,}2)(6{,}1)\} \\ 8 & if\ \omega \in \{(3{,}5)(5{,}3)(4{,}4)(2{,}6)(6{,}2)\} \\ 9 & if\ \omega \in \{(3{,}6)(4{,}5)(5{,}4)(6{,}3)\} \\ 10 & if\ \omega \in \{(4{,}6)(5{,}5)(6{,}4)\} \\ 11 & if\ \omega \in \{(5{,}6)(6{,}5)\} \\ 12 & if\ \omega \in \{(6{,}6)\} \end{cases}$$

ω means Outcome

$$\begin{aligned} \{X=2\} &= \{(1{,}1)\} \\ \{X=3\} &= \{(1{,}2)(2{,}1)\} \\ \{X=4\} &= \{(1{,}3)(2{,}2)(3{,}1)\} \\ \{X=5\} &= \cdots \\ \{X=6\} &= \cdots \\ \{X=7\} &= \cdots \\ \{X=8\} &= \cdots \\ \{X=9\} &= \cdots \\ \{X=10\} &= \{(4{,}6)(5{,}5)(6{,}4)\} \\ \{X=11\} &= \{(5{,}6)(6{,}5)\} \\ \{X=12\} &= \{(6{,}6)\} \end{aligned}$$

Figure: The RV X is defined as the sum of points (x, y)

Discrete & Continuous Random Variable

$$\xrightarrow{\textit{Random Variables}} X, Y, Z, W$$

$$\xrightarrow{\textit{the values of random variables}} x, y, z, w$$

Types:

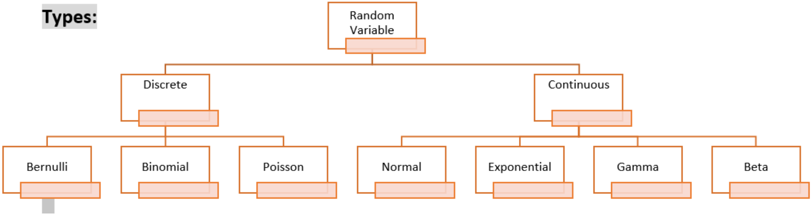


Figure: Some well-known Random Variables

PDF, PMF, CDF

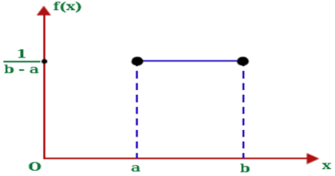
Discrete Random Variable		Continuous Random Variable	
Probability distribution of X		**Probability density function**	
$\mathbf{f(x) = P(X = x)}$		$\mathbf{f(x)}$	
A function can serve as the probability distribution function of a discrete random variable X if and only if its values, f(x), satisfy the condition.	1) $\mathbf{f(x) \geq 0}$ for each value within its domain; 2) $\sum_x \boldsymbol{f(x)} = \mathbf{1}$, where the summation extends over all the values within its domain.	A function can serve as the probability density function of a continuous random variable X if and only if its values, f(x), satisfy the condition.	1) $\mathbf{f(x) \geq 0}$ for each value within its domain; 2) $\int_{-\infty}^{\infty} \boldsymbol{f(x)dx} = \mathbf{1}$, where the summation extends over all the values within its domain.
Distribution function of X (or cumulative distribution of X (CDF))		**Distribution function of X (or cumulative distribution of X (CDF))**	
$\mathbf{F(x) = P(X \leq x)} = \sum_{y \leq x} \mathbf{P(X = y)}$		$\mathbf{F(x) = P(X \leq x)} = \int_{-\infty}^{x} \boldsymbol{f(x)dx}$ $\boldsymbol{for} - \infty < \boldsymbol{x} < \infty$	

PDF, CDF: Example

A random variable X has a Uniform Distribution and it is referred to as a continuous uniform random variable if and only if its probability density is given by

PDF: $f(x)$

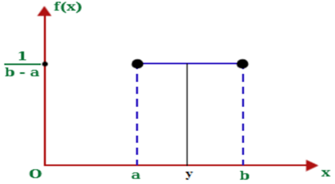
$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha}, & \text{for } \alpha < x < \beta \\ 0, & \text{elsewhere.} \end{cases}$$



Note: $f(x)$ is a density function, because

$$\int_{\alpha}^{\beta} f(x) \cdot dx = \int_{\alpha}^{\beta} \frac{1}{\beta - \alpha} \cdot dx = \frac{1}{\beta - \alpha} \cdot \int_{\alpha}^{\beta} dx$$
$$= \frac{1}{\beta - \alpha} \cdot (\beta - \alpha) = 1$$

X~Unif(α, β)



Note: If $X \sim \text{Unif}(\alpha, \beta)$ then

$$F(y) = \mathrm{P}(X \leq y) = \int_{\alpha}^{y} f(x) \cdot dx$$
$$= \int_{\alpha}^{y} \frac{1}{\beta - \alpha} \cdot dx = \frac{(y - \alpha)}{\beta - \alpha}$$

Derivative formulas

$\frac{d}{dx}(c) = 0$	$\frac{d}{dx}(x) = 1$	$\frac{d}{dx}(cx) = c \cdot \frac{d}{dx}(x) = c$
$\frac{d}{dx}(u + v + \cdots) = \frac{d}{dx}(u) + \frac{d}{dx}(v) + \cdots$		
$\frac{d}{dx}(uv) = u \cdot \frac{d}{dx}(v) + v \cdot \frac{d}{dx}(u)$	$\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{v \cdot \frac{d}{dx}(u) - u \cdot \frac{d}{dx}(v)}{v^2}, \qquad v \neq 0$	
$\frac{d}{dx}(uvw) = uv \cdot \frac{d}{dx}(vw) + uw \cdot \frac{d}{dx}(v) + vw \cdot \frac{d}{dx}(u)$		
$\frac{d}{dx}(cu) = c \cdot \frac{d}{dx}(u), \quad \frac{d}{dx}\left(\frac{c}{u}\right) = c \cdot \frac{d}{dx}\left(\frac{1}{u}\right) = -\frac{c}{u^2} \cdot \frac{d}{dx}(u), \qquad u \neq 0$		
$\frac{d}{dx}(x^m) = m \cdot x^{m-1}$		
$\frac{d}{dx}(u^m) = m \cdot u^{m-1} \frac{d}{dx}(u)$	$z = y^m$ $y = u(x)$	
$\frac{d}{dx}(e^x) = e^x \;, \; \frac{d}{dx} \ln x = \frac{1}{x}$	$\frac{d}{dx} e^{g(x)} = g'(x) \cdot e^{g(x)}$	
$\frac{d}{dx} a^x = a^x \cdot \ln(a), \; \frac{d}{dx} \ln[g(x)] = \frac{g'(x)}{g(x)}$		
$\frac{d}{dx} f[g(x)] = g(x)' \cdot f'[g(x)]$		

Integration formulas

$\int a\,dx = a \cdot x$	$\int a\,f(x)\,dx = a \int f(x)\,dx$
$\int (u \pm v)\,dx = \int u\,dx \pm \int v\ \ dx$	$\int u\,dv = u \cdot v - \int v\,du$
$\int x^n\,dx = \frac{x^{n+1}}{n+1} \qquad (\mathrm{n} \neq 1)$	$\int \frac{dx}{x} = ln\|x\|$
$\int \sqrt{x^m}\ \ dx = \frac{2x \cdot \sqrt{x^m}}{m+2} \qquad (\mathrm{m} \neq -2)$	$\int \sqrt[p]{m}\ \ dx = \frac{px \cdot \sqrt[p]{x^m}}{m+p} \qquad (\mathrm{m} + \mathrm{p} \neq 0)$
$\int e^x\,dx = e^x$	$\int e^{ax}\,dx = \frac{e^{ax}}{a} \qquad (a \neq 0)$

Continuous Random Variable
If X is a continuous random variable and f(x) is the value of its probability density at x, the expected value of X is $$\mathrm{E(X)} = \int_{-\infty}^{\infty} x \cdot f(x)dx$$
If X is a continuous random variable and f(x) is the value of its probability density at x, the expected value of g(X) is given by $$\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f(x)dx$$

The mean and variance of the uniform distribution X are:

$$E(X) = \int_{\alpha}^{\beta} x \cdot f(x) \cdot dx = \int_{\alpha}^{\beta} x \cdot \frac{1}{\beta-\alpha} \cdot dx = \frac{1}{\beta-\alpha} \cdot \int_{\alpha}^{\beta} x \cdot dx = \frac{1}{\beta-\alpha} \cdot \frac{x^2}{2}\Big|_{\alpha}^{\beta} = \frac{\alpha+\beta}{2}$$

$$E[X^2] = \int_{\alpha}^{\beta} x^2 \cdot f(x) \cdot dx = \int_{\alpha}^{\beta} x^2 \cdot \frac{1}{\beta-\alpha} \cdot dx = \frac{1}{\beta-\alpha} \cdot \int_{\alpha}^{\beta} x^2 \cdot dx = \frac{1}{\beta-\alpha} \cdot \frac{x^3}{3}\Big|_{\alpha}^{\beta}$$

$$= \frac{1}{3(\beta-\alpha)} \cdot (\beta^3 - \alpha^3)$$

and the variance is following by

$$\mathrm{Var}(X) = E[X^2] - [E(X)]^2 = \frac{(\beta^3-\alpha^3)}{3(\beta-\alpha)} - \left(\frac{\alpha+\beta}{2}\right)^2 = \frac{1}{12}(\beta-\alpha)^2$$

Multiple Random Variables

<table>
<tr><th colspan="2">Joint PMF: Two RVs X & Y.</th><th colspan="2">Joint PDF: Two RVs X & Y.</th></tr>
<tr><td colspan="2">$f_{X,Y}(x, y) = P(X \le x, Y \le y)$</td><td colspan="2">$f_{X,Y}(x, y) = P(X \le x, Y \le y)$</td></tr>
<tr><td>A bivariate function can serve as the joint probability density function of a pair of discrete random variables X and Y if and only if its values, f(x, y), satisfy the conditions</td><td>1) $f(x,y) \ge 0$ for each pair of values (x, y) within its domain;
2) $\sum_x \sum_y f(x,y) = 1$, where the double summation extends over all possible pairs (x, y) within its domain.</td><td>A bivariate function can serve as the joint probability density function of a pair of continues random variables X and Y if and only if its values, f(x, y), satisfy the conditions</td><td>1) $f(x,y) \ge 0$ for each pair of values (x, y) within its domain;
2) $\int_{-\infty}^{y} \int_{-\infty}^{x} f(x,y)dx \; dy = 1$, where the double summation extends over all possible pairs (x, y) within its domain.</td></tr>
<tr><th colspan="2">Joint CDF: Two RVs X & Y.</th><th colspan="2">Joint CDF: Two RVs X & Y.</th></tr>
<tr><td colspan="2">$$F_{X,Y}(x,y) = P(X \le x, Y \le y) = \sum_{s \le x} \sum_{t \le y} f(s,t)$$ For $-\infty < x < \infty, -\infty < y < \infty$</td><td colspan="2">$$F_{X,Y}(x,y) = P(X \le x, Y \le y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s,t)ds \; dt$$ For $-\infty < x < \infty, -\infty < y < \infty$</td></tr>
<tr><th colspan="2">Marginal Distribution:</th><th colspan="2">Marginal Density Function:</th></tr>
<tr><td colspan="2">If X and Y are discrete random variables and f(x, y) is the value of their joint probability distribution at (x, y), the function given by</td><td colspan="2">If X and Y are continuous random variables and f(x, y) is the value of their joint probability density at (x, y), the function given by</td></tr>
<tr><td>For each x within the range of X is called the marginal distribution of X.</td><td>$$P(X = x) = \sum_{y} P(X = x, Y = y)$$</td><td>$$f_X(x) = \int_{-\infty}^{\infty} f(x,y) \, dy$$ $for \; -\infty < x < \infty$</td><td>$f_X(x)$ is called the marginal density of X.</td></tr>
</table>

Let X and Y be continuous random variables having joint probability density function:

$$f_{X,Y}(x,y) = \begin{cases} k(x^2+y^2), & 0 \le x \le 1, \quad 0 \le y \le 1 \\ 0, & \text{O.W} \end{cases}$$

a) Find the constant k.

b) Compute $P\left(X > \frac{1}{2}, Y < \frac{1}{2}\right)$.

c) Are X and Y independent? Justify your answer.

a)	Verify $\int_0^1 \int_0^1 f_{X,Y}(x,y)\,dx\,dy = 1$
	$\int_0^1 \int_0^1 f_{X,Y}(x,y)\,dx\,dy = \int_0^1 \int_1^1 k(x^2+y^2)\,dx\,dy = \int_0^1 \int_1^1 (kx^2+ky^2)\,dx\,dy = \int_0^1 \left[\int_1^1 (kx^2+ky^2)\,dx\right] dy$ $= \int_0^1 \left[\int_0^1 kx^2\,dx + \int_0^1 \frac{6}{5}y^2\,dx\right] dy = \int_0^1 \left[k \cdot \frac{x^3}{3}\Big\|_0^1 + k \cdot y^2 \cdot x\Big\|_0^1\right] dy = \int_0^1 \left[k \cdot \frac{1}{3} + k \cdot y^2\right] dy$ $= \int_0^1 k \cdot \frac{1}{3}\,dy + \int_0^1 k \cdot y^2\,dy = k \cdot \frac{1}{3} \cdot y\Big\|_0^1 + k \cdot \frac{y^3}{3}\Big\|_0^1 = \frac{1}{3}k + \frac{1}{3}k = \frac{2}{3}k = 1$ $\therefore k = \frac{3}{2}$

b)

$$\int_0^{\frac{1}{2}}\int_{\frac{1}{2}}^{1} f_{X,Y}(x,y)\,dx\,dy = \int_0^{\frac{1}{2}}\int_{\frac{1}{2}}^{1} k(x^2+y^2)\,dx\,dy = \int_0^{\frac{1}{2}}\int_{\frac{1}{2}}^{1}(kx^2+ky^2)\,dx\,dy = \int_0^{\frac{1}{2}}\left[\int_{\frac{1}{2}}^{1}(kx^2+ky^2)\,dx\right]dy$$

$$= \int_0^{\frac{1}{2}}\left[\int_{\frac{1}{2}}^{1} kx^2\ dx + \int_{\frac{1}{2}}^{1} k\,y^2\,dx\right]dy = \int_0^{1}\left[k\cdot\frac{x^3}{3}\Big|_{\frac{1}{2}}^{1} + k\cdot y^2\cdot x\Big|_{\frac{1}{2}}^{1}\right]dy$$

$$= \int_0^{\frac{1}{2}}\left[\frac{1}{3}k - \frac{1}{24}k + k\cdot y^2 - \frac{1}{2}k\cdot y^2\right]dy = 1/4$$

$f_X(x) = \int_0^1 f_{X,Y}(x,y)\,dy$ $= \int_0^1 (kx^2+ky^2)\,dy$ $= \int_0^1 \left[\frac{3}{2}x^2+\frac{3}{2}y^2\right]dy$ $= \frac{3}{2}x^2+\frac{1}{2}$	$f_Y(y) = \int_0^1 f_{X,Y}(x,y)\,dx = \int_0^1 (kx^2+ky^2)\,dx$ $= \int_0^1 \left[\frac{3}{2}x^2+\frac{3}{2}y^2\right]dx$ $= \frac{1}{2}+\frac{3}{2}y^2$

$$f_X(x)\cdot f_Y(y) = \left(\frac{3}{2}x^2+\frac{1}{2}\right)\cdot\left(\frac{1}{2}+\frac{3}{2}y^2\right) = \frac{9}{4}x^2y^2+\frac{1}{4}x^2+\frac{3}{4}y^2+\frac{1}{4} \neq f_{X,Y}(x,y)$$

So the random variable X and Y are not independent to each other.

Some General Advice

- Be confident.
- Be in contact with faculty members.
- Be in touch with your classmates.
- Be organized in your academic life.