# Probability & Statistics Exam 2 Note Sheets

## Chapter 3: Continuous Distributions

**Definitions:**

- **pdf of $X$:** $f \colon S \to \mathbb{R}$ has the following properties:
  - $f(x) \geq 0$ for all $x \in S$; $\int_S f(x)\,dx = 1$; if $(a,b) \subseteq S$, then $P(a \leq X \leq b) = \int_a^b f(x)\,dx$

- **CDF of $X$:** $F(x) = P(X \leq x) = \int_{-\infty}^x f(t)\,dt$

- **Expected Value of $X$:** $E(X) = \mu = \int_S x f(x)\,dx$

- **Variance of $X$:** $\operatorname{Var}(X) = \sigma^2 = \int_S (x - \mu)^2 f(x)\,dx = E(X^2) - \mu^2$

- **mfg of $X$:** $M_X(t) = E(e^{tX}) = \int_S e^{tx} f(x)\,dx$

- **Percentiles:** The $p$th percentile $\pi_p$ is the value such that $P(X \leq \pi_p) = p$. Thus, $\int_{-\infty}^{\pi_p} f(x)\,dx = p$.

- **Integration by Parts:** $\int u\,dv = uv - \int v\,du$.

- **Gamma Function:** $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\,dt$, for $\alpha > 0$.
  - $\Gamma(x) = (x-1)\Gamma(x-1)$, $\Gamma(\tfrac{1}{2}) = \sqrt{\pi}$, $\Gamma(x) = \frac{\Gamma(x+1)}{x}$

**Examples:**

1. $\int_0^1 2x e^{-x^2}\,dx \implies u = x^2,\ du = 2x\,dx \implies \int_0^1 e^{-u}\,du = 1 - e^{-1}$

2. Consider the function $f(x) = cx^5$ on $[0,1]$.
   - (a) Find $c$. (Solve $\int_0^1 cx^5\,dx = 1$.)
   - (b) Determine $E(X)$. (Solve $\int_0^1 x \cdot cx^5\,dx$.)
   - (c) Determine $\operatorname{Var}(X)$. (Solve $\int_0^1 (x - \mu)^2 \cdot cx^5\,dx$.)
   - (d) Determine the 25th percentile. (Solve $\int_0^{\pi_{25}} cx^5\,dx = 0.25$.)

3. A pdf is given by $f(x) = \tfrac{1}{2}$. for $x \in [0,1] \cup [2,3]$, and 0 otherwise.
   - (a) Determine $E(X)$. (Solve $\int_0^1 x \cdot \tfrac{1}{2}\,dx + \int_2^3 x \cdot \tfrac{1}{2}\,dx$. To do the variance, just replace $x$ with $x^2$ and then subtract $\mu^2$.)

4. The mean airspeed of an unladen swallow is 30 ft/s, with a std. dev. of 4.3 ft/s. Assuming the airspeed is normally distributed:
   - (a) What is the probability that a randomly selected swallow has an airspeed between 28 and 38 ft/s? ($P(28 \leq X \leq 38) = $ `normalcdf(28, 38, 30, 4.3)`)

## Chapter 3 (cont.)

4. (b) Suppose that 12 swallows are selected at random. What is the probability that exactly 8 have their speed between 28 and 38 ft/s? (Use binomial formula with $n = 12$, $k = 8$, and $p$ from previous problem.)

5. An auto insurance company insures an automobile worth \$15,000 for one year under a policy with a \$1,000 deductible. During the policy year there is a 0.04 chance of partial damage to the car and a 0.02 chance of a total loss of the car. If there is partial damage to the car, the amount $X$ of damage (in thousands) follows a distribution with density function $f(x) = 0.5003 e^{-x/2}$, for $0 \leq x \leq 15$. Find the expected value of the payment the insurance company makes in a year.

   *Solution:* Let $E$ = expected yearly payment. There are 3 cases: 1. No damage: $P = 0.94$, payment $= 0$. 2. Total loss: $P = 0.02$, payment $= 14{,}000$. 3. Partial damage: $P = 0.04$, $X \sim f(x) = 0.5003 e^{-x/2}$, $0 \leq x \leq 15$ (in thousands). The payment for $X > 1$ is $1000(X - 1)$ and 0 otherwise. Expected payout for partial damage is $\int_1^{15} 1000(X-1)0.5003 e^{-x/2}\,dx = 1211.96$. Thus, $E = 0.94 \cdot 0 + 0.02 \cdot 14000 + 0.04 \cdot 1211.96 = 328.48$.

## Chapter 4: Bivariate Distributions

**Definitions:**

- **Joint pmf:** $f(x,y) = P(X = x, Y = y)$

- **Marginal pmf of $X$:** $f_X(x) = \sum_y f(x,y) = P(X = x)$ for $x \in S_X$

- **Mean of $X_i$:** $\mu_{X_i} = E(X_i) = \sum_{x_i} x_i f_{X_i}(x_i)$

- **Variance of $X_i$:** $\operatorname{Var}(X_i) = \sigma_{X_i}^2 = \sum_{x_i} (x_i - \mu_{X_i})^2 f_{X_i}(x_i) = E(X_i^2) - \mu_{X_i}^2$

- **Covariance of $X$, $Y$:** $\operatorname{Cov}(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X \mu_Y$. If $\operatorname{Cov}(X,Y) = 0$, then $X$ and $Y$ are uncorrelated; if pos, $X+$ and $Y+$, if neg, $X+$ and $Y-$. If independent, then $\operatorname{Cov}(X,Y) = 0$.

- **Correlation of $X$ and $Y$:** $\rho_{X,Y} = \frac{\operatorname{Cov}(X,Y)}{\sigma_X \sigma_Y}$, or $\rho = 0$ if independent.

- **Conditional pmf of $X$ given $Y$:** $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$ for $y \in S_Y$

- **Conditional Expectation of $X$ given $Y$:** $E(X|Y = y) = \sum_x x f_{X|Y}(x|y)$

- **Conditional Variance of $X$ given $Y$:** $\operatorname{Var}(X|Y = y) = E(X^2|Y = y) - (E(X|Y = y))^2$

- **Least Squares Regression Line:** $y = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(X - \mu_x)$

- **Independence Criteria:** If $f(x,y) = f_X(x) f_Y(y)$ for all $x \in S_X$, $y \in S_Y$.

**Examples:**

1. An actuary determines that the annual number of tornadoes in counties $P$ and $Q$ are jointly distributed as follows:

|  | County $Q$ | | | |
|---|---|---|---|---|
| County $P$ | 0 | 1 | 2 | 3 |
| 0 | 0.12 | 0.06 | 0.05 | 0.02 |
| 1 | 0.13 | 0.15 | 0.12 | 0.03 |
| 2 | 0.05 | 0.15 | 0.10 | 0.02 |

   (a) Determine the conditional expected number of tornados in county $Q$, given that there are no tornados in county $P$.
   $(Q|P = 0) = 0 \cdot \frac{0.12}{0.25} + 1 \cdot \frac{0.06}{0.25} + 2 \cdot \frac{0.05}{0.25} + 3 \cdot \frac{0.02}{0.25} = 0.88.$

   (b) Calculate the conditional variance of the annual number of tornadoes in county $Q$, given that there are no tornados in county $P$.
   $\text{Var}(Q|P = 0) = E(Q^2|P = 0) - (E(Q|P = 0))^2 = 1.76 - 0.88^2 = 0.9856.$

   (c) Are counties $P$ and $Q$ independent? Justify your answer.
   No, because $f(0, 0) = 0.12 \neq f_P(0)f_Q(0) = 0.25 \cdot 0.30 = 0.075.$

2. Let $X$ be the weight of robin eggs, in grams, and $Y$ be the daily high temperature, in degrees Celsius. Assume that $X$ and $Y$ have a bivariate normal distribution with $\mu_X = 145.2$, $\sigma_X^2 = 109.2$, $\mu_Y = 23.5$, $\sigma_Y^2 = 21.8$ and $\rho = -0.34$.

   (a) Find the least squares regression line for predicting $Y$ from $X$.
   $y = 23.5 + (-0.34)\frac{\sqrt{21.8}}{\sqrt{109.2}}(X - 145.2) = 23.5 - 0.152(X - 145.2)$

   (b) Using the line from part (a), predict the daily high temperature when the weight of a robin egg is 150 grams.
   $y = 23.5 - 0.152(150 - 145.2) = 22.76$

   (c) Find the probability that a robin egg weights between 142 and 152 grams.
   $P(142 \leq X \leq 152) = \texttt{normalcdf(142, 152, 145.2, 10.45)}$

   (d) Given $Y = 25.1$, find the probability that a robin egg weighs between 142 and 152 grams. (Use CDF formula w updated $\mu$ and $\sigma$:)
   $\mu_{X|Y=25.1} = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(25.1 - \mu_Y) = 143.9824$
   $\sigma_{X|Y=25.1}^2 = (1 - \rho^2)\sigma_X^2 = 96.5765 \implies \sigma_{X|Y=25.1} = \sqrt{96.5765} = 9.82$

# Chapter 5: Distributions of Functions of Random Variables

The formula for a z-score is $z = \frac{x-\mu}{\sigma}$. The formula for converting a z-score to an x-value is $x = \mu + z\sigma$.

**Examples:**

1. Let $X$ be a random variable with pdf $f(x) = cx^3$ for $0 < x < 2$. Let $Y = X^4$. Determine the pdf of $Y$. (cdf $= F(y) = P(Y \leq y) = P(X^4 \leq y) = P(X \leq y^{1/4}) = F_X(y^{1/4})$, then differentiate to get pdf.)

2. Let $X$ and $Y$ be independent random variables with $f(x) = 2x, 0 < x < 1$ and $g(y) = 4y^3, 0 < y < 1$. Find $(0.5 < X < 1$ and $0.4 < Y < 0.8)$. (Since they are independent, just multiply their individual probabilities.)

3. Three basketball players each make 70% of their free throws. They each take 10 free throws. Assuming independence among the players, what is the probability that at least one player makes more than 8? (This is the complement of all 3 making 8 or fewer. Hence, $1 - (\texttt{binomcdf(10, 0.7, 8)})^3 = 0.3844$.)

4. Suppose that each of $Z_1, Z_2, \ldots, Z_9$ are distributed as $N(0, 1)$. Let $W = Z_1^2 + Z_2^2 + \ldots + Z_9^2$. Find $P(7.2 < W < 16)$. (Since $W$ is chi-square with 9 degrees of freedom, use $\texttt{x2cdf(7.2, 16, 9)}$.)

5. 30 students roll a single 4-sided die until they get exactly 5 ones. Let $\overline{X}$ be the average number of rolls needed. Use the CLT to approximate $P(17 < \overline{X} < 19.5)$. (The number of rolls needed follows a negative binomial distribution with $p = \frac{1}{4}$ and $r = 5$. Thus, $\mu_X = \frac{r}{p} = 20$ and $\sigma_X^2 = \frac{r(1-p)}{p^2} = 60 \implies \sigma_X = \sqrt{60} = 7.746$. By the CLT, $\overline{X} \sim N(20, \frac{60}{30}) = N(20, 2)$. Hence, $P(17 < \overline{X} < 19.5) = \texttt{normalcdf(17, 19.5, 20, sqrt(2))} = 0.3449$.)

6. A certain exam has a mean score of 73 and std. dev. of 6.8. Use CLT to approximate the probability that $n = 19$ students who took the exam have a mean larger than 75. (By CLT, $\overline{X} \sim N(73, \frac{6.8^2}{19}) = N(73, 2.434)$. Thus, $P(\overline{X} > 75) = \texttt{normalcdf(75, 1E99, 73, sqrt(2.434))} = 0.0999$.)

7. A certain brand of LED lightbulbs advertise that they have a mean lifetime of 12.4 years, and we will assume their lifetime is exponentially distributed. You purchase a random sample of size $n = 34$, and find that the mean lifetime of your sample is 9.5 years. Use the Central Limit Theorem to (quantitatively) comment on how likely or unlikely this is? [Hint: Since the normal distribution is continuous, it does not make sense to talk about the probability that a sample has a particular mean. It might make sense to talk about the proportion of samples whose mean is at most some particular number, however.]
   We know that an individual bulb has a lifetime with $\mu = 12.4$ and $\sigma^2 = 12.4^2 = 153.76$, assuming that the company's claim is correct. The mean lifetime of our sample would then follow $N(12.4, 153.76/34)$. Thus, the probability of getting a sample with your mean lifetime or even shorter is given by $\texttt{normalcdf(-1e99, 9.5, 12.4, 12.4/sqrt(34))} = 0.08633$. This is reasonably unlikely, either you got a fairly unusual sample or the company's claim about the mean lifetime is incorrect.