Stat 230: Formulas and Review

Combinatorics

1. $n^{(k)} = \frac{n!}{(n-k)!}$
2. $\binom{n}{k} = \frac{n^{(k)}}{k!}$

Sets and Probability:

1. $\overline{A \cup B} = \overline{A} \cap \overline{B}$
2. $\overline{A \cap B} = \overline{A} \cup \overline{B}$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
4. $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$
5. $P(A) = 1 - P(\bar{A})$
6. **Independent events:** $P(A \cap B) = P(A)P(B)$
7. **Conditional probability** $P(A|B) = \frac{P(A \cap B)}{P(B)}$ provided $P(B) > 0$.
8. **Product Rules:**
   (a) $P(A \cap B) = P(A)P(B|A)$
   (b) $P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$
9. **Law of Total Probability** *If* $A_1 \cup A_2 \cup \cdots \cup A_k = S$ and $A_i \cap A_j = \emptyset$ if $i \neq j$ then for B *an arbitrary event*

$$P(B) = P(BA_1) + P(BA_2) + \cdots + P(BA_k) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

10. **Review of Series**
    (a) **Geometric Series:** $a + ar + ar^2 + \cdots + ar^{n-1} = \frac{a(1-r^n)}{1-r}$ for $r \neq 1$
    (b) **Binomial Theorem:** $(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}$
    (c) **Exponential Series:** $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$

$f(x) = F(x) - F(x-1)$

Random Variables and distributions

1. **Cumulative distribution function** (c.d.f.) of X: $F(x) = P(X \leq x)$, for all $x \in \Re$. $\lim_{x \to -\infty} F(x) = 0$ $\lim_{x \to \infty} F(x) = 1$

2. **Various distributions** -see summary of distributions Notes page 341 (provided on exam)

3. **Expected value** of a discrete random variable:

   (a) $E(X) = \sum_{\text{all x}} xf(x)$ (denoted μ)

   (b) $E[ag(X) + b] = aE[g(X)] + b$ for any function g.

4. **Variance** of a random variable X:

   (a) $Var(X) = E\left[(X - \mu)^2\right] = E(X^2) - \mu^2$

   (b) **Standard deviation:** $\sigma = \sqrt{Var(X)}$

5. **Properties of Mean and Variance:** If $Y = aX + b$, then $E(Y) = aE(X) + b$ and $Var(Y) = a^2 Var(X)$.

Continuous Distributions

1. $P(a < X < b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = F(b) - F(a)$

2. **Probability density function** (p.d.f.) $f(x) = \frac{dF(x)}{dx}$.

   (a) $P(a \leq X \leq b) = \int_a^b f(x)dx$.

   (b) $f(x) \geq 0$.

   (c) $\int_{-\infty}^{\infty} f(x)dx = \int_{\text{all}x} f(x)dx = 1$.

   (d) $F(x) = \int_{-\infty}^{x} f(u)du$.

3. **Expectation (Mean), and Variance**

   (a) $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$

   (b) $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$

   (c) $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$

(d) $Var(X) = E[(X-\mu)^2] = E(X^2) - \mu^2$

4. **Normal distribution** $N(\mu, \sigma^2)$:

   (a) $EX) = \mu$, $Var(X) = \sigma^2$:

   (b) **p.d.f.** $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$ for $x \in \Re$

   (c) **Standardizing:** If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

Multivariate Distributions

1. **Joint probability function of** (X, Y):
   $f(x, y) = P(X = x \text{ and } Y = y) = P(X = x, Y = y)$.

2. **marginal probability function of** X: $f_1(x) = f_x(x) = \sum_{\text{all } y} f(x, y)$

3. **marginal probability function of** Y: $f_2(y) = f_y(y) = \sum_{\text{all } x} f(x, y)$

4. **Independent** random variables X and Y: $f(x, y) = f_1(x)f_2(y)$ for all (x, y)

5. **conditional probability function** of X given $Y = y$ is $f_1(x|y) = \frac{f(x,y)}{f_2(y)}$ provided $f_2(y) > 0$.

6. **Sum of Independent Poisson:** If $X \backsim Poisson\,(\mu_1)$ and $Y \backsim Poisson\,(\mu_2)$ independently then $X + Y \backsim Poisson\,(\mu_1 + \mu_2)$.

7. **Multinomial Joint Probability Function:** $f(x_1, x_2, \ldots, x_k) = \frac{n!}{x_1!x_2!\cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$.

8. **Expected value:** $E[g(X, Y)] = \sum_{\text{all } (x,y)} g(x, y)f(x, y)$.

   (a) **Example:** $E(XY) = \sum_{\text{all } (x,y)} xyf(x, y)$

   (b) **Linearity of Expected Value:** $E[ag_1(X, Y) + bg_2(X, Y)] = aE[g_1(X, Y)] + bE[g_2(X, Y)]$ and so $E(aX + bY) = aE(X) + bE(Y)$

9. **Covariance, Variance, Correlation:**

   (a) $Cov(X, Y) = E[(X-\mu_x)(Y-\mu_y)] = E(XY) - \mu_x\mu_y$, where $\mu_x = E(X), \mu_y = E(Y)$.

   (b) If X and Y are **independent** and g_1 and g_2 are any two functions, $E[g_1(X)g_2(Y)] = E[g_1(X)]E[g_2(Y)]$.

   (c) If X and Y are **independent** then $Cov(X, Y) = 0$

   (d) **Correlation Coefficient of** X **and** Y**:** $\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$

(e) $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$

(f) For a, b, c, and d constants, $\mathrm{Cov}(aX + bY, cU + dV) = ac\mathrm{Cov}(X, U) + ad\mathrm{Cov}(X, V) + bc\mathrm{Cov}(Y, U) + bd\mathrm{Cov}(Y, V)$.

(g) $Var\,(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab\mathrm{Cov}(X, Y)$

(h) If $X_1, X_2, \cdots, X_n$ are independent, $Var\left(\sum\limits_{i=1}^{n} a_i X_i\right) = \sum\limits_{i=1}^{n} a_i^2 Var(X_i)$.

10. **Linear Combination of Normal:** If $X_i \sim N\left(\mu_i, \sigma_i^2\right)$, $i = 1, 2, \ldots, n$ independently, then $\sum\limits_{i=1}^{n} a_i X_i \sim N\left(\sum\limits_{i=1}^{n} a_i \mu_i, \sum\limits_{i=1}^{n} a_i^2 \sigma_i^2\right)$ (for any n)

Normal Approximations and Moment Generating Functions

1. **Central Limit Theorem** If $X_1, X_2, \ldots, X_n$ independent with the same distribution, then as $n \to \infty$, the c.d.f. of the random variable $\frac{\sum_{i=1}^{n} X_i - n\mu_x}{\sigma_x \sqrt{n}} = \frac{\overline{X} - \mu_x}{\sigma_x / \sqrt{n}}$ approaches the $N(0, 1)$ c.d.f., where $\overline{X} = \frac{1}{n} \sum\limits_{i=1}^{n} X_i$.

   (a) **Normal Approximation to Poisson:** Suppose $X \sim Poisson(\mu)$. Then $Z = \frac{X - \mu}{\sqrt{\mu}}$ has approximately a $N(0, 1)$ distribution for μ large.

   (b) **Normal Approximation to Binomial:** Suppose $X \backsim Binomial(n, p)$. Then the random variable $W = \frac{X - np}{\sqrt{np(1-p)}}$ has approximately a $N(0, 1)$ distribution for n large.

2. **Moment generating function** (m.g.f.) of X is $M(t) = E(e^{tX})$ defined for all t such that this is finite.