MA224: Probability Test 2 Outline

Sections covered: 2.6, 3.1, 3.3-6, 4.1, some of 4.2, some of 5.1, 5.3-5.7.

2.6. The Poisson Distribution

- Def. 2.6.1. Let the number of changes that occur in a given continuous interval be counted. Then we have an approximate Poisson distribution with parameter $\lambda > 0$ if the conditions are satisfied:
  - The numbers of changes occurring in nonoverlapping intervals are independent
  - Probability of exactly one change occurring in short time interval h is λh
  - Probability of two or more changes in short time interval is 0
- Poisson distribution is binomial distribution $b\left(n, \dfrac{\lambda}{n}\right)$
  - Do n trials, so average is λ
  - $\lim\limits_{n\to\infty} f(x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$
    - For a time interval t, distribution is $f(x) = \dfrac{(\lambda t)^x e^{-\lambda t}}{x!}$
  - $\mu = M'(0) = \lambda = \sigma^2$
- Can approximate binomial distribution with high n, p (and therefore λ) should be small
  - $\dfrac{(np)^x e^{-np}}{x!} \approx \dbinom{n}{x} p^x (1-p)^{n-x} = b(n,p)$ (replace λ with np)
  - In turn can be used to approximate hypergeometric distribution with high n and small p (i.e., $N_1 = np$, $N_2 = n - np$)

3.1. Continuous-Type Data, 3.2. Exploratory Data Analysis

- (not useful)

3.3. Random Variables of the Continuous Type

- Probability density function (pdf) of a continuous random variable X is an integratable function such that $P(a < X < b) = \displaystyle\int_a^b f(x)\,dx$
  - $f(x) = 0$ when $x \notin S$
- Cumulative density function is $F(x) = P(X \le x) = \displaystyle\int_{-\infty}^{x} f(x)\,dx$
  - $P(X = b) = 0$
  - $P(a \le X \le b) = F(b) - F(a)$
- Same other definitions that were based on pdfs for discrete-type variables:
  - $\mu = E(x) = \displaystyle\int_{-\infty}^{\infty} x f(x)\,dx$
  - $\sigma^2 = E\left[(X-\mu)^2\right] = \displaystyle\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$
  - $M(t) = E\left(e^{tx}\right) = \displaystyle\int_{-\infty}^{\infty} e^{tx} f(x)\,dx,\ -h < t < h$
- (From exercise 12) For function $R(t) = \ln M(t)$:
  - $\mu = R'(0)$
  - $\sigma^2 = R''(0)$
- Median found by using cdf: $F(x) = 0.5$

3.4. The Uniform and Exponential Distributions

- Uniform distribution:
  - $F(x) = \dfrac{x-a}{b-a},\ a \le x \le b$
  - $f(x) = \dfrac{1}{b-a},\ a \le x \le b$
  - $M(t) = \dfrac{e^{bt} - e^{at}}{t(b-a)}$
  - $\mu = \dfrac{a+b}{2}$
  - $\sigma^2 = \dfrac{(b-a)^2}{12}$
- Exponential distribution:
  - $F(w) = 1 - e^{-\lambda w} = 1 - e^{-\frac{w}{\theta}}$
    - $\Rightarrow P(X \ge x) = e^{-\frac{x}{\theta}}$
    - $\Rightarrow m\,(\text{median}) = \theta \ln 2$
  - $M(t) = \dfrac{1}{1-\theta t},\ t < \dfrac{1}{\theta}$
  - $f(w) = \lambda e^{-\lambda w} = \dfrac{1}{\theta} e^{-\frac{w}{\theta}},\ \theta = \dfrac{1}{\lambda},\ \theta > 0$
  - $\mu = \theta$, $\sigma^2 = \theta^2 \Rightarrow \sigma = \theta$
  - i.e., while λ is average changes per unit time, θ is average waiting time between changes (makes sense; should be inversely proportional)
  - "Failure rate is constant": i.e., conditional probability over the same change in time is constant, i.e., $P(X > t_1 + t | X > t_1) = P(X > t)\ (= (P(X > t | X > 0)))$; has real-world application that its not worth replacing an object with constant failure rate
    - Similar to (discrete) geometric distribution; no other continuous distribution has this "forgetfulness" property

3.5. The Gamma and Chi-Square Distributions

- Gamma distribution:
  - Let W be the time until the α-th change occurs
  - $F(w) = \dfrac{\lambda(\lambda w)^{\alpha-1}}{(a-1)!} e^{-\lambda w}$
    - If $w < 0$, $F(w) = 0$, $F'(w) = 0$
  - Definition of gamma function, useful in writing the pdf of the distribution: $\Gamma(t) = \displaystyle\int_0^\infty y^{t-1} e^{-y}\, dy = (t-1)!,\ t > 0$
  - $f(x) = \dfrac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-\frac{x}{\theta}},\ 0 \le x < \infty$
  - $M(t) = \dfrac{1}{(1-\theta t)^\alpha},\ t < \dfrac{1}{\theta}$
  - $\mu = \alpha\theta$, $\sigma^2 = \alpha\theta^2$
- Chi-square distribution: Gamma distribution with $\theta = 2$, $\alpha = \dfrac{r}{2}$, r is a positive integer
  - $f(x) = \chi^2(r) = \dfrac{1}{\Gamma(\frac{r}{2}) 2^{\frac{r}{2}}} x^{\frac{r}{2}-1} e^{-\frac{x}{2}}$
  - $P(x \ge \chi^2_\alpha(r)) = \alpha$ (use lookup table for this)
  - $\mu = r,\ \sigma^2 = 2r$

- $M(t) = \dfrac{1}{(1-2t)^{-\frac{r}{2}}},\ t < \dfrac{1}{2}$
  - Exponential distribution with $\mu = \theta = 2$ is $\chi^2(2)$

3.6. The Normal Distribution

- $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} \exp\left[-\dfrac{(x-\mu)^2}{2\sigma^2}\right]\ dx$
- $M(t) = \exp\left(\mu t + \dfrac{\sigma^2 t^2}{2}\right)$
- μ and σ^2 are given in the pdf
- $N(0,1)$ is called the standard normal distribution
- Sometimes want to find inverse of standard normal distribution function
  - I.e., normally find $P(X \le x) = p$, but now find x given p (use lookup tables)
  - Also sometimes want to find $P(z \ge z_\alpha) = \alpha$ (find z_α, the upper percent)
  - $P(Z \le -z_\alpha) = P(Z \ge z_\alpha) = \alpha$ (because of symmetry)
  - $z_{1-\alpha} = -z_\alpha$
  - Think of z_α as the z-score of the normal distribution where α of the distribution is greater than that z-scores
- Theorem 3.6-1. Standardizing a normal distribution. If X is $N(\mu, \sigma^2)$, then $Z = \dfrac{X-\mu}{\sigma} = N(0,1)$
  - (Basically taking z-score of distribution)
- Theorem 3.6-2. If the random variable X is $N(\mu, \sigma^2),\ \sigma^2 > 0$, then the random variable $V = \dfrac{(X-\mu)^2}{\sigma^2} = Z^2$ is $\chi^2(1)$.

4.1. Distributions of Two Random Variables

- Definition 4.1-1. Joint pmf $f(x,y) = P(X = x, Y = y)$
  - $\sum\sum_{(x,y)\in S} f(x,y) = 1$ (do a double summation instead of a single one)
- Marginal pmf of X is $f_1(x) = \sum_y f(x,y) = P(X = x),\ x \in S_1$
  - i.e., sum it over the other axis to add up all of its variability
  - X and Y are independent IFF $f(x,y) = f_1(x) f_2(x)$
    - Independence can only happen if the support is “rectangular”
- Mathematical expectation:
  - $\mu_x = E(u(x,y)) = E(x)$
  - $\sigma_x^2 = E(u(x,y)) = E((x-\mu)^2)$
  - μ_x and σ_x^2 can be calculated either from joint pmf or marginal pmf
- Continuous analogue with integrals
- Examples:
  - Hypergeometric pmf of multiple variables; marginal pmfs are also hypergeometric and are dependent
  - Binomial pmf to trinomial distribution pmf: marginal pmfs are also binomial and dependent

4.2. The Correlation Coefficient

- If $u(X_1, X_2) = (X_1 - \mu_1)(X_2 - \mu_2)$, $E(u(X_1, X_2)) = \sigma_{12}$ (covariance of X_1 and X_2)
- $\rho = \dfrac{\sigma_{12}}{\sigma_1 \sigma_2}$ is called the correlation coefficient (if the stdevs are positive)

- Need the joint pmf to compute (not a marginal one)
- $\sigma_{12} = E(X_1 X_2)$
- $E(X_1 X_2) = \mu_1 \mu_2 + \rho \sigma_1 \sigma_2$

5.1. Functions of One Random Variable

- Two methods to find the distribution of a function of a random variable (also a random variable)
  - Distribution function technique
    - Plug in the function into the cdf of the second variable
    - e.g., if X is a random variable, and $Y = u(X)$, express $P(Y \leq y) = P(u(X) \leq y) = P(X \leq u^{-1}(y))$ **,** and plug into the cdf of X, and then differentiate to get the pdf of Y
    - e.g., loggamma is the substitution of $Y = e^X$ if X is the gamma distribution, $g(y) = \dfrac{1}{\Gamma(\alpha)\theta^\alpha} \dfrac{(\ln y)^{\alpha-1}}{y^{1+\frac{1}{\theta}}}$
      - $\mu = \dfrac{1}{(1-\theta)^\alpha}$, $\sigma^2 = \dfrac{1}{(1-2\theta)^\alpha} - \dfrac{1}{(1-\theta)^{2\alpha}}$
    - e.g., Cauchy pdf (p. 217)
  - Change of variable technique: shortcut for distribution function technique (same methodology)
    - $g(y) = f\,[v(y)]\,|v'(y)\,|$

5.3. Several Independent Variables

- Dealing with the pmf resulting from repeated mutually-independent trials (i.e., each marginal pmf/pdf is the same)
- If all n distributions are the same, then called random sample of size n from that common distribution, and $g(x_1, x_2, \cdots, x_n) = f(x_1) f(x_2) \cdots f(x_n)$
- Theorem 5.3-1. Generalization of expected value (discrete form) $E(Y) = \sum_y y g(y) = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} u(x_1, x_2, \cdots, x_n) f_1(x_1) f_2(x_2) \cdots f_n(x_n)$
- Theorem 5.3-2. If $Y = u_1(X_1) u_2(X_2) \cdots u_n(X_n)$, $E(Y) = E(u_1(X_1)) E(u_2(X_2)) \cdots E(u(X_n))$
- If $Y = \sum_{i=1}^{n} a_i X_i$, then $\mu_Y = \sum_{i=1}^{n} a_i \mu_i$ and $\sigma_Y^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2$
- For a random sample of n samples, where $\bar{X}$ is a function of the samples, $\mu_{\bar{X}} = \mu$, $\sigma_{\bar{X}}^2 = \dfrac{\sigma^2}{n}$

5.4. The Moment-Generating Function Technique

- Theorem 5.4-1. The moment-generating function $Y = \sum_{i=1}^{n} a_i X_i$ is $M_Y(t) = \prod_{i=1}^{n} M_{x_i}(a_i t)$
  - Corollary: for $Y = \sum X_i$, $M_Y(t) = [M(t)]^n$
  - Corollary: for $\bar{X} = \sum \dfrac{1}{n} X_i$, $M_{\bar{X}}(t) = \left[M\left(\dfrac{t}{n}\right)\right]^n$
- Examples:
  - For a bernoulli trial, $M(t) = q + pe^t$. For $Y = \sum X_i$, get $M_Y(t) = \left(q + pe^t\right)^n$ $(b(n,p))$
  - For a Chi-square trial, Y is sum of trials, get $\chi^2\left(\sum_i r_i\right)$

- If samples of normal standard distributions ($N(0,1)$), and $W = \sum_{i=1}^{n} Z_i^2$, then $W = \chi^2(n)$ (using theorem 3.6-2)
- If samples $X_i = N(\mu_i, \sigma_i^2)$, then $W = \sum_{i=1}^{n} \frac{(X_i - \mu_i)^2}{\sigma_i^2} = \chi^2(n)$ (using theorem 3.6-1 and above corollary)

5.5. Random Functions Associated with Normal Distributions

- Theorem 5.5-1. If random samples X_i are independent normal distributions, then $Y = \sum_{i=1}^{n} c_i X_i$ has the distribution $N\left(\sum_{i=1}^{n} c_i \mu_i, \sum_{i=1}^{n} c_i^2 \sigma_i^2\right)$
  - Corollary: distribution of $\bar{X}$ is $N\left(\mu, \frac{\sigma^2}{n}\right)$

5.6. The Central Limit Theorem

- Theorem 5.6-1. Central Limit Theorem. The distribution of $\lim_{n\to\infty} W = \lim_{n\to\infty} \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \lim_{n\to\infty} \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$ is $N(0,1)$. (W is a function of $\bar{X}$, which is in turn a function of X).
  - $P(W \leq w) \approx \int_{-\infty}^{w} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz = \phi(w)$
    - e.g., for $X = N(\mu, \sigma)$ approximation for $P(a < \bar{X} < b) \approx P\left(\frac{a - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{b - \mu}{\frac{\sigma}{\sqrt{n}}}\right)$
- CLT useful when distribution is symmetric, unimodal, and continuous
  - General rule of thumb is that it is a good approximation for large n ($n > 25$), but can be smaller if unimodal and symmetric
- For a single distribution of X, $W = \frac{X - \mu}{\sigma}$ can be used to create a Normal distribution to approximate X
  - $G(x) = F(\sigma x + \mu)$, $g(x) = \sigma f(\sigma x + \mu)$

5.7. Approximations for Discrete Distributions

- … Working here …