# Review for Midterm II

## §2.1 Distribution of Two Random Variables

- Two-dimensional random vector $(X_1, X_2)$:
  $X_1$ and $X_2$ are random variables defined on the same sample space $\mathcal{C}$.
- Space of $(X_1, X_2)$: $\mathcal{D} = \{(X_1(c), X_2(c)) : c \in \mathcal{C}\}$
- Joint cdf of $(X_1, X_2)$: $F_{X_1,X_2}(x_1, x_2) = P(X_1 \le x_1, X_2 \le x_2)$
- MGF of $(X_1, X_2)$: $M(t_1, t_2) = E[e^{t_1X_1+t_2X_2}]$
- For discrete $(X_1, X_2)$:
  - ∗ Joint pmf: $p_{X_1,X_2}(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$
  - ∗ Marginal pmf: $p_1(x_1) = \sum_{x_2} p_{X_1,X_2}(x_1, x_2)$
    $p_2(x_2) = \sum_{x_1} p_{X_1,X_2}(x_1, x_2)$
  - ∗ Probability of event $A$: $P(A) = \sum\sum_{(x_1,x_2)\in A} p_{X_1,X_2}(x_1, x_2)$
  - ∗ Expectation: $E[g(X_1, X_2)] = \sum_{x_1}\sum_{x_2} g(x_1, x_2)p_{X_1,X_2}(x_1, x_2)$
- For continuous random vector $(X_1, X_2)$:
  - ∗ Joint pdf: $f_{X_1,X_2}(x_1, x_2)$, such that for all $x_1, x_2 \in R$,
    $F_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_{X_1,X_2}(w_1, w_2)dw_2dw_1$
  - ∗ Marginal pdf: $f_1(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)dx_2$
    $f_2(x_2) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)dx_1$
  - ∗ Probability of event $A$: $P(A) = \int\int_{(x_1,x_2)\in A} f_{X_1,X_2}(x_1, x_2)dx_1dx_2$
  - ∗ Expectation: $E[g(X_1, X_2)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1, x_2)f_{X_1,X_2}(x_1, x_2)dx_1dx_2$
- Property of expectations:
  $E[k_1g_1(X_1, X_2) + k_2g_2(X_1, X_2)] = k_1E[g_1(X_1, X_2)] + k_2E[g_2(X_1, X_2)]$
- For practice: Example 2.1.3, Example 2.1.4, Example 2.1.5, Example 2.1.7

## §2.2 Transformations: Bivariate Random Variables

- Let $(X_1, X_2)$ be a random vector with support $\mathcal{S}$.
  One-to-one transformation mapping $\mathcal{S}$ onto $\mathcal{T}$:
  $Y_1 = u_1(X_1, X_2)$, $Y_2 = u_2(X_1, X_2)$
  Inverse transformation mapping $\mathcal{T}$ onto $\mathcal{S}$:
  $X_1 = w_1(Y_1, Y_2)$, $X_2 = w_2(Y_1, Y_2)$
- For discrete $(X_1, X_2)$, the joint pmf of $(Y_1, Y_2)$:
  $p_{Y_1,Y_2}(y_1, y_2) = p_{X_1,X_2}[w_1(y_1, y_2), w_2(y_1, y_2)]$, $(y_1, y_2) \in \mathcal{T}$

- For continuous $(X_1, X_2)$, the joint pdf of $(Y_1, Y_2)$:
  $f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}[w_1(y_1, y_2), w_2(y_1, y_2)] \cdot |J|$, $(y_1, y_2) \in \mathcal{T}$,
  where $J$ is the Jacobian determinant of the transformation

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \frac{\partial x_1}{\partial y_1} \cdot \frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_2} \cdot \frac{\partial x_2}{\partial y_1}$$

- For practice: Example 2.2.1, Example 2.2.4, Example 2.2.5

## §2.3 Conditional Distributions and Expectations

- For discrete $(X_1, X_2)$, the conditional pmf's are:
  $p_{2|1}(x_2|x_1) = p(x_1, x_2)/p_1(x_1)$, $p_{1|2}(x_1|x_2) = p(x_1, x_2)/p_2(x_2)$
  Conditional probability: $P(X_2 \in A|X_1 = x_1) = \sum_{x_2 \in A} p_{2|1}(x_2|x_1)$
  Conditional expectation: $E[u(X_2)|X_1 = x_1] = \sum_{x_2} u(x_2) p_{2|1}(x_2|x_1)$
- For continuous $(X_1, X_2)$, the conditional pdf's are:
  $f_{2|1}(x_2|x_1) = f(x_1, x_2)/f_1(x_1)$, $f_{1|2}(x_1|x_2) = f(x_1, x_2)/p_2(x_2)$
  Conditional probability: $P(X_2 \in A|X_1 = x_1) = \int_{x_2 \in A} f_{2|1}(x_2|x_1) dx_2$
  Conditional expectation: $E[u(X_2)|X_1 = x_1] = \int_{-\infty}^{\infty} u(x_2) f_{2|1}(x_2|x_1) dx_2$
- Conditional variance: $\mathrm{Var}(X_2|x_1) = E(X_2^2|x_1) - [E(X_2|x_1)]^2$
- Conditional expectation given $X_1$: $E(X_2|X_1)$, a function of $X_1$
  Properties: $E[E(X_2|X_1)] = E(X_2)$, $\mathrm{Var}[E(X_2|X_1)] \leq \mathrm{Var}(X_2)$
- For practice: Example 2.3.1, Example 2.3.2

## §2.4 Covariance and Correlation Coefficient

- Let $(X, Y)$ be a random vector. Denote
  $\mu_1 = E(X)$, $\mu_2 = E(Y)$, $\sigma_1^2 = \mathrm{Var}(X)$, $\sigma_2^2 = \mathrm{Var}(Y)$
- Covariance of $X$ and $Y$:
  $\mathrm{Cov}(X, Y) = E[(X - \mu_1)(Y - \mu_2)] = E(XY) - \mu_1\mu_2$
- Correlation coefficient of $X$ and $Y$: $\rho = \mathrm{Cov}(X, Y)/(\sigma_1\sigma_2)$
- Properties of conditional expectation:
  If $E(Y|X)$ is linear in $X$, then $E(Y|X) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1)$ .
  If $E(X|Y)$ is linear in $Y$, then $E(X|Y) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(Y - \mu_2)$ .
- For practice: Example 2.4.1, Example 2.4.2

## §2.5 Independent Random Variables

- Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be the supports of $X_1$ and $X_2$ respectively.
- For discrete $(X_1, X_2)$, $X_1$ and $X_2$ is said to be independent if and only if one of the following conditions is true:
  [1] the joint pmf $p(x_1, x_2) = p_1(x_1)p_2(x_2)$ for all $x_1$ and $x_2$, where $p_1$, $p_2$ are

marginal pmf's of $X_1$ and $X_2$ respectively;

[2] the joint pmf $p(x_1, x_2) = g(x_1)h(x_2)$ for all $x_1$ and $x_2$, where $g(x_1) > 0$ for $x_1 \in \mathcal{S}_1$ and $h(x_2) > 0$ for $x_2 \in \mathcal{S}_2$ .

- For continuous $(X_1, X_2)$, $X_1$ and $X_2$ is said to be independent if and only if one of the following conditions is true:

  [1] the joint pdf $f(x_1, x_2) = f_1(x_1)f_2(x_2)$ for all $x_1$ and $x_2$, where $f_1$, $f_2$ are marginal pdf's of $X_1$ and $X_2$ respectively;

  [2] the joint pdf $f(x_1, x_2) = g(x_1)h(x_2)$ for all $x_1$ and $x_2$, where $g(x_1) > 0$ for $x_1 \in \mathcal{S}_1$ and $h(x_2) > 0$ for $x_2 \in \mathcal{S}_2$ .

- Other equivalent definitions: $X_1$ and $X_2$ is said to be independent if and only if one of the following conditions is true:

  [3] the joint cdf $F(x_1, x_2) = F_1(x_1)F_2(x_2)$ for all $x_1$ and $x_2$ , where $F_1$ and $F_2$ are the margianl cdf of $X_1$ and $X_2$;

  [4] $P(a < X_1 \leq b, c < X_2 \leq d) = P(a < X_1 \leq b)P(c < X_2 \leq d)$ for all $a, b, c, d$, such that $a < b$ and $c < d$;

  [5] Joint MGF $M(t_1, t_2) = M(t_1, 0)M(0, t_2)$ for all $-h < t_1, t_2 < h$ .

- Other properties: If $X_1$ and $X_2$ are independent, then
  [6] $E[u(X_1)v(X_2)] = E[u(X_1)] \cdot E[v(X_2)]$;
  [7] $\operatorname{Cov}(X_1, X_2) = 0$, $\rho(X_1, X_2) = 0$;
  [8] $\operatorname{Var}(X_1 + X_2) = \operatorname{Var}(X_1) + \operatorname{Var}(X_2)$

- For practice: Example 2.5.1, Example 2.5.2

## §2.6 Extension to Several Random Variables

- $\mathbf{X} = (X_1, \ldots, X_n)$ is called $n$-dimensional random vector if $X_1, \ldots, X_n$ are random variables defined on the same sample space $\mathcal{C}$.
- Space of $\mathbf{X}$: $\mathcal{D} = \{(X_1(c), \ldots, X_n(c)) : c \in \mathcal{C}\}$
- Joint cdf of $\mathbf{X}$: $F_{\mathbf{X}}(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$
- MGF of $\mathbf{X}$: $M(t_1, \ldots, t_n) = E[e^{t_1X_1+\cdots+t_nX_n}]$
- For discrete random vector $(X_1, \cdots, X_n)$:
  - ∗ Joint pmf: $p_{\mathbf{X}}(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$
  - ∗ Marginal pmf: $p_1(x_1) = \sum_{x_2} \cdots \sum_{x_n} p_{\mathbf{X}}(x_1, \ldots, x_n)$
    $p_{1,2}(x_1, x_2) = \sum_{x_3} \cdots \sum_{x_n} p_{\mathbf{X}}(x_1, \ldots, x_n)$
  - ∗ Joint conditional pmf:
    $$p_{2,\ldots,n|1}(x_2, \ldots, x_n|x_1) = p_{\mathbf{X}}(x_1, \ldots, x_n)/p_1(x_1)$$
  - ∗ Probability of event $A$: $P(A) = \sum_{(x_1,\ldots,x_n)\in A} p_{\mathbf{X}}(x_1, \ldots, x_n)$
  - ∗ Expectation:
    $$E[g(X_1, \ldots, X_n)] = \sum_{x_1} \cdots \sum_{x_n} g(x_1, \ldots, x_n)p_{\mathbf{X}}(x_1, \ldots, x_n)$$
  - ∗ $X_1, \ldots, X_n$ are said to be independent if and only if one of the following conditions is true:

[1] the joint pmf $p_{\mathbf{X}}(x_1, \ldots, x_n) = p_1(x_1) \cdots p_n(x_n)$ for all $x_1, \ldots, x_n$;

[2] the joint pmf $p_{\mathbf{X}}(x_1, \ldots, x_n) = g_1(x_1) \cdots g_n(x_n)$ for all $x_1, \ldots, x_n$, where $g_i(x_i) > 0$ for $x_i \in \mathcal{S}_i$, $i = 1, \ldots, n$ .

- For continuous random vector $(X_1, \cdots, X_n)$:
  - $*$ Joint pdf: $f_{\mathbf{X}}(x_1, \ldots, x_n)$, such that for all $x_1, \ldots, x_n \in R$,
    $$F_{\mathbf{X}}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathbf{X}}(w_1, \ldots, w_n) dw_n \cdots dw_1$$
  - $*$ Marginal pdf: $f_1(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, \ldots, x_n) dx_2 \cdots dx_n$
    $f_{1,2}(x_1, x_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, \ldots, x_n) dx_3 \cdots dx_n$
  - $*$ Joint conditional pmf:
    $$f_{2,\ldots,n|1}(x_2, \ldots, x_n | x_1) = f_{\mathbf{X}}(x_1, \ldots, x_n) / f_1(x_1)$$
  - $*$ Probability of event $A$: $P(A) = \int_{(x_1,\ldots,x_n) \in A} f_{\mathbf{X}}(x_1, \ldots, x_n) dx_1 \cdots dx_n$
  - $*$ Expectation:
    $E[g(X_1, \ldots, X_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f_{\mathbf{X}}(x_1, \ldots, x_n) dx_1 \cdots dx_n$
  - $*$ $X_1, \ldots, X_n$ are said to be independent if and only if one of the following conditions is true:

    [1] the joint pdf $f_{\mathbf{X}}(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n)$ for all $x_1, \ldots, x_n$;

    [2] the joint pdf $f_{\mathbf{X}}(x_1, \ldots, x_n) = g_1(x_1) \cdots g_n(x_n)$ for all $x_1, \ldots, x_n$, where $g_i(x_i) > 0$ for $x_i \in \mathcal{S}_i$, $i = 1, \ldots, n$ .
- Other equivalent definitions: $X_1, \ldots, X_n$ are said to be mutually independent if and only if one of the following conditions is true:

  [3] the joint cdf $F_{\mathbf{X}}(x_1, \ldots, x_n) = F_1(x_1) \cdots F_n(x_n)$ for all $x_1, \ldots, x_n$;

  [4] $P(a_1 < X_1 \le b_1, \ldots, a_n < X_n \le b_n) = P(a_1 < X_1 \le b_1) \cdots P(a_n < X_n \le b_n)$ for all $a_i < b_i$, $i = 1, \ldots, n$;

  [5] the joint MGF $M(t_1, \ldots, t_n) = M(t_1, 0, \ldots, 0) \cdots M(0, \ldots, 0, t_n)$ for all $-h < t_i < h$, $i = 1, \ldots, n$ .
- Other properties: If $X_1, \ldots, X_n$ are mutually independent, then
  [6] $u_1(X_1), \ldots, u_n(X_n)$ are independent;
  [7] Any $k$ random variables of $X_1, \ldots, X_n$ are independent, $2 \le k \le n$;
  [8] $E[u_1(X_1) \cdots u_n(X_n)] = E[u_1(X_1)] \cdots E[u_n(X_n)]$;
  [9] $\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)$
- Random variables $X_1, \ldots, X_n$ are said to be independent and identically distributed (i.i.d.), if they are independent and have the same distribution.
- For practice: Example 2.6.2, Remark 2.6.1

## §2.7 Transformations: Random Vectors

- Let $(X_1, \ldots, X_n)$ be a random vector with support $\mathcal{S}$.
  One-to-one transformation mapping $\mathcal{S}$ onto $\mathcal{T}$:
  $Y_1 = u_1(X_1, \ldots, X_n),\ \ldots,\ Y_n = u_n(X_1, \ldots, X_n)$
  Inverse transformation mapping $\mathcal{T}$ onto $\mathcal{S}$:
  $X_1 = w_1(Y_1, \ldots, Y_n),\ \ldots,\ X_n = w_n(Y_1, \ldots, Y_n)$

- In discrete case, the joint pmf of $(Y_1, \ldots, Y_n)$:
  $p_{\mathbf{Y}}(y_1, \ldots, y_n) = p_{\mathbf{X}}[w_1(y_1, \ldots, y_n), \ldots, w_n(y_1, \ldots, y_n)]$, $(y_1, \ldots, y_n) \in \mathcal{T}$
- In continuous case, the joint pdf of $\mathbf{Y} = (Y_1, \ldots, Y_n)$:
  $f_{\mathbf{Y}}(y_1, \ldots, y_n) = f_{\mathbf{X}}[w_1(y_1, \ldots, y_n), \ldots, w_n(y_1, \ldots, y_n)] \cdot |J|$, $(y_1, \ldots, y_n) \in \mathcal{T}$, where $J$, the Jacobian determinant of the transformation, is

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}$$

- For practice: Example 2.7.1, Example 2.7.2

## §3.1 The Binomial and Related Distributions

- Bernoulli trial: $\mathcal{C} = \{\text{success}, \text{failure}\}$, $P(\{\text{success}\}) = p = 1 - P(\{\text{failure}\})$
- Bernoulli distribution with parameter $p$ (success rate), $0 < p < 1$:

$$p(x) = p^x (1-p)^{1-x}, x = 0, 1$$

  Mean and Variance: $\mu = p$, $\sigma^2 = p(1-p)$
  MGF: $M(t) = (1-p) + pe^t$, $t \in R$
- Binomial distribution with parameters $n$ and $p$, $0 < p < 1$:
  Let $X$ be the number of successes in $n$ independent Bernoulli trials, then $X \sim B(n, p)$ with pmf

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, 2, \ldots, n$$

  Mean and Variance: $\mu = np$, $\sigma^2 = np(1-p)$
  MGF: $M(t) = [(1-p) + pe^t]^n$
  If $X_1, \ldots, X_n$ are i.i.d. $\sim$ Bernoulli($p$), then $Y = X_1 + \cdots + X_n \sim B(n, p)$ .
  If $X_1, \ldots, X_m$ are independent and $X_i \sim B(n_i, p)$, $i = 1, \ldots, m$, then $Y = X_1 + \cdots + X_m$ has $B(\sum_{i=1}^m n_i, p)$ distribution.
- Negative binomial distribution with parameters $r$ and $p$, $0 < p < 1$:
  Let $Y$ be the number of failures before the $r$th success in a sequence of independent Bernoulli trials, then $Y \sim NB(r, p)$:

$$p(y) = \binom{y+r-1}{r-1} p^r (1-p)^y, y = 0, 1, 2, \ldots$$

  Mean and Variance: $\mu = \frac{r(1-p)}{p}$, $\sigma^2 = \frac{r(1-p)}{p^2}$
  MGF: $M(t) = p^r[1 - (1-p)e^t]^{-r}$, $t < -\log(1-p)$
  Note: $NB(r = 1, p)$ is the geometric distribution.

- Multinomial distribution with parameters $n$ and $p_1, \ldots, p_{k-1}$:
  Repeat a random experiment with sample space $\{1, 2, \ldots, k\}$ for $n$ independent times. Let $X_i$ be the number of outcome $i$, $i = 1, \ldots, k$. Then $(X_1, \ldots, X_{k-1})$ has Multinomial$(n;\, p_1, \ldots, p_{k-1})$ distribution.

  $$p(x_1, \ldots, x_{k-1}) = \frac{n!}{x_1! \cdots x_{k-1}! x_k!} p_1^{x_1} \cdots p_{k-1}^{x_{k-1}} p_k^{x_k},$$

  where Let $p_k = 1 - (p_1 + \cdots + p_{k-1})$, and $x_k = n - (x_1 + \cdots + x_{k-1})$.
  MGF: $M(t_1, \ldots, t_{k-1}) = (p_1 e^{t_1} + \cdots + p_{k-1} e^{t_{k-1}} + p_k)^n$
  If $(X_1, X_2) \sim$ Multinomial$(n; p_1, p_2)$ , then its marginal distribution: $X_1 \sim B(n, p_1)$, $X_2 \sim B(n, p_2)$.
- For practice: Example 3.1.3, Example 3.1.5

## §3.2 The Poisson Distribution

- Poisson distribution with rate parameter $\lambda > 0$, denoted by Poisson$(\lambda)$ :

  $$p(x) = e^{-\lambda} \frac{\lambda^x}{x!}, x = 0, 1, 2, \ldots$$

- Mean and Variance: $\mu = \sigma^2 = \lambda$
- MGF: $M(t) = e^{\lambda(e^t - 1)}$
- If $X_1, \ldots, X_n$ are independent and $X_i \sim$ Poisson$(\lambda_i)$, then $\sum_{i=1}^n X_i \sim$ Poisson$(\sum_{i=1}^n \lambda_i)$.
- For practice: Example 3.2.1, Example 3.2.3

## §3.4 The Normal Distribution

- Normal distribution with mean $\mu$ and variance $\sigma^2$, denoted by $N(\mu, \sigma^2)$

  $$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right\}, -\infty < x < \infty$$

- MGF: $M(t) = \exp\left\{ \mu t + \frac{1}{2}\sigma^2 t^2 \right\}$, $-\infty < t < \infty$
- Transformation: If $X \sim N(\mu, \sigma^2)$, then $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$.
  Standardization: $Z = (X - \mu)/\sigma \sim N(0, 1)$
- Standard normal distribution: $N(0, 1)$ with Pdf and cdf:

  $$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad \Phi(z) = \int_{-\infty}^{z} \phi(x) dx$$

  Property of $\Phi$:$\Phi(-z) = 1 - \Phi(z)$
- If $X_1, \ldots, X_n$ are i.i.d. $\sim N(0, 1)$, then $Y = \sum_{i=1}^n X_i^2 \sim \chi^2(n)$ .
- If $X_1, \ldots, X_n$ are independent and $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$, then

  $$Y = \sum_{i=1}^{n} a_i X_i \sim N(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2).$$

- For practice: Example 3.4.3, Example 3.4.4