Today: Probability review (cont.)

Next: PK 2.1, 2.3

This week:

- all office hours are on course web site

- HW1 due Friday, January 17, 23:59 pm
Joint normal distribution (cont.)

Last time:

Proposition 1. Let \( \mathbf{\mu} \in \mathbb{R}^2 \), let \( \mathbf{\Sigma} \in \mathbb{R}^{2 \times 2} \) be positive definite, let \( \mathbf{Y} = (Y_1, Y_2)^t \) be independent standard Gaussian r.v.'s. Let \( \mathbf{A} \in \mathbb{R}^{2 \times 2} \) be such that \( \mathbf{\Sigma} = \mathbf{AA}^T \) (always possible).

Then \( \mathbf{A} \mathbf{Y} + \mathbf{\mu} \sim \mathcal{N}(\mathbf{\mu}, \mathbf{\Sigma}) \).

Corollary 1. Let \( \mathbf{X} \sim \mathcal{N}(\mathbf{\mu}, \mathbf{\Sigma}) \), \( \mathbf{\Sigma} = \mathbf{AA}^T \). Then

Corollary 2. Any linear combination of Gaussian random vectors is again a Gaussian random vector.
Joint normal distribution (cont.)

**Corollary 3** If \( \mathbf{X} = (X_1, X_2) \sim \mathcal{N}(\mathbf{\mu}, \Sigma) \), then

\[
\text{Proof. Compute } \mathbb{E}((\mathbf{X} - \mathbf{\mu}) (\mathbf{X} - \mathbf{\mu})^t) =
\]

Let \( Y_1, Y_2 \) be independent \( \mathcal{N}(0,1) \), denote \( \mathbf{Y} = (Y_1, Y_2) \), and let \( \mathbf{A} \) be such that \( \Sigma = \mathbf{A} \mathbf{A}^t \).

Then
Compute the density explicitly: let \( \mathbf{\mu} = (\mu_1, \mu_2) \in \mathbb{R}^2 \),
\[
\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} > 0,
\]
let \( \mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}(\mathbf{\mu}, \mathbf{\Sigma}) \).

\[
f_{X}(x) = \frac{1}{2\pi \sqrt{\det \mathbf{\Sigma}}} e^{-(x-\mathbf{\mu})^t \mathbf{\Sigma}^{-1} (x-\mathbf{\mu})}
\]

\[
\det \mathbf{\Sigma} = \sigma_1 \sigma_2
\]

Then, \( \sqrt{\det \mathbf{\Sigma}} = \sigma_1 \sigma_2 \)

Similarly, \( \mathbf{\Sigma}^{-1} = \)

\[
(x-\mathbf{\mu})^t \mathbf{\Sigma}^{-1} (x-\mathbf{\mu}) = \]
Remarks 1) Definition 1, Proposition 1 and its corollaries 1-3 are not restricted to random vectors of dimension 2: we can generalize the definition by taking \( \tilde{\mu} \in \mathbb{R}^r, \Sigma \in \mathbb{R}^{r \times r}, \Sigma > 0, \) to define a random Gaussian vector \( \tilde{X} \in \mathbb{R}^r, \tilde{X} \sim \mathcal{N}(\tilde{\mu}, \Sigma) \) with the joint p.d.f.

\[
f_{\tilde{X}}(x) =
\]

Proposition 1 and Corollaries 1-3 remain unchanged.
Joint normal distribution (cont.)

2) The joint normal distribution is uniquely identified by the

\[ \Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \]

Ex. We have 3 r.v. \( X, Y, Z \)

\[ X \sim \text{Unif}[-1, 1], \quad Y \sim \text{Unif}[-1, 1], \quad Z \sim \text{Unif}[-1, 1] \]

3) If normal random variables are uncorrelated (\( \Sigma = I \)), then
Estimating tail probabilities

Markov Thm. Let $X$ be a r.v., $P(X \geq 0) = 1$ ($X \geq 0$ almost surely).
Then for any $c > 0$

Chebyshev Thm. Let $X$ be a r.v., $E(X) = \mu$, $\text{Var}(X) = \sigma^2$. Then for any $c > 0$

Immediate corollaries

\[ P(X - \mu > c) \leq \]
\[ P(X - \mu \leq -c) \leq \]
**Moment generating function**

**Def.** Let $X$ be a r.v. Then the function

$$M(t) = \mathbb{E}(e^{tX})$$

is called the moment generating function of $X$.

M.g.f. is used to characterize the distribution of a r.v.:

**Thm.** Let $X$ and $Y$ be two r.v.'s, let $M_X(t)$ and $M_Y(t)$ be their m.g.f.'s. If there exists $\delta > 0$ s.t.

(i) $M_X(t)$ and $M_Y(t)$ are bounded

(ii) for all $t \in (-\delta, \delta)$,

then

Computing moments from m.g.f.: if $M_X(t)$ is bounded around $t=0$, then,
Conditional distribution (discrete case)

Recall: for two events $A, B \in \mathcal{F}$, conditional probability of $A$ given $B$ is computed via

\[
\text{Def. Let } X, Y \text{ be two discrete r.v.'s taking values in } \\
\{x_1, x_2, ... \} \text{ and } \{y_1, y_2, ... \} \text{ correspondingly. The conditional probability mass function of } X \text{ given } Y \text{ is defined by }
\]

By the law of total probability ($\{Y=y_j\}_{j=1}^{\infty}$ is a partition)
Conditional distribution. Example
Conditional expectation (discrete case)

Def. Let $X, Y$ be discrete r.v.'s with values $\{x_i\}, \{y_j\}$. Let $g: \mathbb{R} \to \mathbb{R}$ be a function such that $|E(g(X))| < \infty$. The conditional expectation of $g(X)$ given $Y = y_j$ is defined by

By the law of total probability