Lecture-12: Characteristic function

1 Transforms for distribution functions

Suppose that $X : \Omega \to \mathbb{R}$ is a random variable on the probability space $(\Omega, \mathcal{F}, P)$ with distribution function $F_X : \mathbb{R} \to [0,1]$.

Example 1.1. A function $g_\theta : \mathbb{R} \to \mathbb{R}$ defined by $g_\theta(x) \triangleq e^{\theta x}$ is Borel measurable for all $\theta \in \mathbb{R}$. Therefore, $g_\theta(X)$ is a positive random variable on this probability space. We can show that $h_\theta : \mathbb{R} \to \mathbb{C}$ defined by $h_\theta(x) \triangleq e^{j\theta x} = \cos(\theta x) + j\sin(\theta x)$ is also Borel measurable for all $\theta \in \mathbb{R}$, where $j = \sqrt{-1}$. Thus, $h_\theta(X)$ is a complex valued random variable on this probability space.

Remark 1. Recall that if $\mathbb{E}|X|^N$ is finite for some $N \in \mathbb{N}$, then $\mathbb{E}|X|^k$ is finite for all $k \in [N]$. This follows from the linearity and monotonicity of expectations, and the fact that

$$|X|^k = |X|^k \mathbb{1}_{\{|X| \leqslant 1\}} + |X|^k \mathbb{1}_{\{|X| > 1\}} \leqslant 1 + |X|^N.$$

1.1 Characteristic function

Definition 1.2 (Characteristic function). For a random variable $X : \Omega \to \mathbb{R}$ defined on the probability space $(\Omega, \mathcal{F}, P)$, the **characteristic function** $\Phi_X : \mathbb{R} \to \mathbb{C}$ is defined by $\Phi_X(u) \triangleq \mathbb{E}e^{juX}$ for all $u \in \mathbb{R}$ and $j^2 = -1$.

Remark 2. Since $e^{j\theta} = \cos\theta + j\sin\theta$ for real $\theta \in \mathbb{R}$, the characteristic function can be equivalently written as $\Phi_X(u) = \mathbb{E}[\cos uX] + j\mathbb{E}[\sin uX]$.

Remark 3. Suppose that $X : \Omega \to \mathcal{X}$ is a discrete random variable with PMF $P_X : \mathcal{X} \to [0,1]$, then $\Phi_X(u) = \sum_{x \in \mathcal{X}} e^{jux} P_X(x)$.

Remark 4. Suppose that $X : \Omega \to \mathbb{R}$ is a continuous random variable with density function $f_X : \mathbb{R} \to \mathbb{R}_+$, then $\Phi_X(u) = \int_{-\infty}^{\infty} e^{juX} f_X(x) dx$.

Remark 5. The characteristic function $\Phi_X(u)$ is always finite, since $|\Phi_X(u)| = \left|\mathbb{E}e^{juX}\right| \leqslant \mathbb{E}\left|e^{juX}\right| = 1$.

Example 1.3 (Gaussian random variable). For a Gaussian random variable $X : \Omega \to \mathbb{R}$ with mean μ and variance σ^2, the characteristic function Φ_X is

$$\Phi_X(u) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{x \in \mathbb{R}} e^{jux} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \exp\Big(-\frac{u^2\sigma^2}{2} + ju\mu\Big).$$

We observe that $|\Phi_X(u)| = e^{-u^2\sigma^2/2}$ has Gaussian decay with zero mean and variance $1/\sigma^2$.

Theorem 1.4. *If $\mathbb{E}|X|^N$ is finite for some integer $N \in \mathbb{N}$, then $\Phi_X^{(k)}(u)$ is finite and continuous functions of u for all $k \in [N]$. Further, $\Phi_X^{(k)}(0) = j^k \mathbb{E}[X^k]$ for all $k \in [N]$.*

Proof. Exchanging derivative and the integration (which can be done since e^{jux} is a bounded function with all derivatives), and evaluating the derivative at $u = 0$, we get

$$\Phi_X^{(k)}(0) = \mathbb{E}\left[\frac{d^k e^{juX}}{du^k}\Big|_{u=0}\right] = j^k \mathbb{E}[X^k].$$

Since $\mathbb{E}|X|^N$ is finite, then so is $\mathbb{E}|X|^k$ for all $k \in [N]$. Therefore, $\mathbb{E}[X^k]$ exists and is finite, and $\Phi_X^{(k)}(0) = j^k \mathbb{E}[X^k]$.

Theorem 1.5. *Two random variables have the same probability distribution iff they have the same characteristic function.*

Proof. It is easy to see the necessity and the sufficiency is difficult. □

1.2 Moment generating function

Characteristic function always exist, however are complex in general. Sometimes it is easier to work with moment generating functions, when they exist.

Definition 1.6 (Moment generating function). For a random variable $X : \Omega \to \mathbb{R}$ defined on the probability space $(\Omega, \mathcal{F}, P)$, the moment generating function $M_X : \mathbb{R} \to \mathbb{R}_+$ is defined by $M_X(t) \triangleq \mathbb{E}e^{tX}$ for all $t \in \mathbb{R}$ where $M_X(t)$ is finite.

Lemma 1.7. *For a random variable X, if the MGF $M_X(t)$ is finite for some $t \in \mathbb{R}$, then $M_X(t) = 1 + \sum_{n\in\mathbb{N}} \frac{t^n}{n!}\mathbb{E}[X^n]$.*

Proof. From the Taylor series expansion of e^θ around $\theta = 0$, we get $e^\theta = 1 + \sum_{n\in\mathbb{N}} \frac{\theta^n}{n!}$. Therefore, taking $\theta = tX$, we can write

$$e^{tX} = 1 + \sum_{n\in\mathbb{N}} \frac{t^n}{n!} X^n.$$

Taking expectation on both sides, the result follows from the linearity of expectation, when both sides have finite expectation.

Example 1.8 (Gaussian random variable). For a Gaussian random variable $X : \Omega \to \mathbb{R}$ with mean μ and variance σ^2, the moment generating function M_X is

$$M_X(t) = \exp\left(\frac{t^2\sigma^2}{2} + t\mu\right).$$

1.3 Probability generating function

For a non-negative integer-valued random variable X it is often more convenient to work with the z-transform of the PMF, called the probability generating function.

Definition 1.9. For a discrete random variable $X : \Omega \to \mathcal{X}$ with probability mass function $P_X : \mathcal{X} \to [0,1]$, the **probability generating function** $\Psi_X : \mathbb{C} \to \mathbb{C}$ is defined by

$$\Psi_X(z) \triangleq \mathbb{E}[z^X] = \sum_{x\in\mathcal{X}} z^x P_X(x), \quad z \in \mathbb{C}, |z| \leqslant 1.$$

Lemma 1.10. *For a simple random variable $X : \Omega \to \mathcal{X}$, we have $|\Psi_X(z)| \leqslant 1$ for all $|z| \leqslant 1$.*

Proof. Let $z \in \mathbb{C}$ with $|z| \leqslant 1$. Let $P_X : \mathcal{X} \to [0,1]$ be the probability mass function of the positive simple random variable X. Since any realization $x \in \mathcal{X}$ of random variable X is positive, we can write

$$|\Psi_X(z)| = \left| \sum_{x \in \mathcal{X}} z^x P_X(x) \right| \leqslant \sum_{x \in \mathcal{X}} |z|^x P_X(x) \leqslant \sum_{x \in \mathcal{X}} P_X(x) = 1.$$

Theorem 1.11. *For a positive simple random variable X, the k-th derivative of probability generating function evaluated at $z = 1$ is the k-th order factorial moment of X. That is,*

$$\Psi_X^{(k)}(1) = \mathbb{E}\left[\prod_{i=0}^{k-1}(X - i)\right] = \mathbb{E}[X(X-1)(X-2)\dots(X-k+1)].$$

Proof. It follows from the interchange of derivative and expectation.

Remark 6. Moments can be recovered from kth order factorial moments. For example,

$$\mathbb{E}[X] = \Psi_X'(1), \qquad \mathbb{E}[X^2] = \Psi_X^{(2)}(1) + \Psi_X'(1).$$

2 Gaussian Random Vectors

Definition 2.1. For a random vector $X : \Omega \to \mathbb{R}^n$ defined on a probability space $(\Omega, \mathcal{F}, P)$, we can define the characteristic function $\Phi_X : \mathbb{R}^n \to \mathbb{C}$ by $\Phi_X(u) \triangleq \mathbb{E}e^{j\langle u, X\rangle}$ where $u \in \mathbb{R}^n$.

Remark 7. If $X : \Omega \to \mathbb{R}^n$ is an independent random vector, then $\Phi_X(u) = \prod_{i=1}^n \Phi_{X_i}(u_i)$ for all $u \in \mathbb{R}^n$.

Definition 2.2 (*I.i.d.* Gaussian random vector). For a probability space $(\Omega, \mathcal{F}, P)$, an *i.i.d.* Gaussian random vector $X : \Omega \to \mathbb{R}^n$ is a continuous random vector defined by its density function

$$f_X(x) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^n \frac{(x_i - \mu_1)^2}{\sigma^2}\right) \text{ for all } x \in \mathbb{R}^n.$$

for some real scalar μ_1 and positive $\sigma^2 \in \mathbb{R}_+$.

Remark 8. For an *i.i.d.* Gaussian random vector with density parametrized by (μ_1, σ^2), the components are *i.i.d.* Gaussian random variables with mean μ_1 and variance σ^2.

Remark 9. The characteristic function Φ_X of an *i.i.d.* Gaussian random vector $X : \Omega \to \mathbb{R}^n$ parametrized by (μ_1, σ^2) is given by

$$\Phi_X(u) = \prod_{i=1}^n \Phi_{X_i}(u_i) = \exp\Big(-\frac{\sigma^2}{2}\sum_{i=1}^n u_i^2 + j\mu_1 \sum_{i=1}^n u_i\Big).$$

Lemma 2.3. *For an* i.i.d. *zero mean unit variance Gaussian vector $Z : \Omega \to \mathbb{R}^n$, vector $\alpha \in \mathbb{R}^n$, and scalar $\mu \in \mathbb{R}$, the affine combination $Y \triangleq \mu + \langle \alpha, Z\rangle$ is a Gaussian random variable.*

Proof. From the linearity of expectation and the fact that Z is a zero mean vector, we get $\mathbb{E}Y = \mu$. Further, from the linearity of expectation and the fact that $\mathbb{E}[ZZ^T] = I$, we get

$$\sigma^2 \triangleq \mathrm{Var}(Y) = \mathbb{E}(Y - \mu)^2 = \sum_{i=1}^n \sum_{k=1}^n \alpha_i \alpha_k \mathbb{E}[Z_i Z_k] = \langle \alpha, \alpha \rangle = \|\alpha\|_2^2.$$

To show that Y is Gaussian, it suffices to show that $\Phi_Y(u) = \exp(-\frac{u^2\sigma^2}{2} + ju\mu)$ for any $u \in \mathbb{R}$. Recall that Z is an independent random vector with individual components being identically zero mean unit variance Gaussian. Therefore, $\Phi_{Z_i}(\theta) = \exp(-\frac{\theta^2}{2})$, and we can compute the characteristic function of Y as

$$\Phi_Y(u) = \mathbb{E}e^{juY} = e^{ju\mu}\mathbb{E}\prod_{i=1}^n e^{ju\alpha_i Z_i} = e^{ju\mu}\prod_{i=1}^n \Phi_{Z_i}(u\alpha_i) = \exp(-\frac{u^2\sigma^2}{2} + ju\mu).$$

□

Definition 2.4 (Gaussian random vector). For a probability space $(\Omega, \mathcal{F}, P)$, a Gaussian random vector $X : \Omega \to \mathbb{R}^n$ can be written as

$$X = \mu + AZ,$$

from some vector $\mu \in \mathbb{R}^n$, matrix $A \in \mathbb{R}^{n\times n}$, and *i.i.d.* Gaussian random vector $Z : \Omega \to \mathbb{R}^n$ with mean 0 and variance 1. We denote the covariance matrix for the Gaussian vector X by $\Sigma \triangleq \mathbb{E}(X-\mu)(X-\mu)^T$.

Remark 10. The components of the Gaussian random vector are Gaussian random variables with mean μ_i and variance $\sum_{k=1}^n A_{i,k}^2 = (AA^T)_{i,i}$, since each component $X_i = \mu_i + \sum_{k=1}^n A_{i,k} Z_k$ is an affine combination of zero mean unit variance *i.i.d.* random variables.

Lemma 2.5. *For a Gaussian random vector $X = \mu + AZ$ for $\mu \in \mathbb{R}^n, A \in \mathbb{R}^{n\times n}$, and* i.i.d. *zero mean unit variance Gaussian random vector Z, the covariance matrix is $\Sigma = AA^T$.*

Proof. We can write $X_i = \mu_i + \sum_{k=1}^n A_{i,k} Z_k$ and we get $\mathbb{E}X_i = \mu_i$ from linearity of expectations and the fact that $\mathbb{E}Z_k = 0$ for all $k \in [n]$. Similarly, the (i,j)th component of covariance matrix is the mean of

$$(X_i - \mu_i)(X_j - \mu_j) = \sum_{\ell=1}^n \sum_{k=1}^n A_{i,k} A_{j,\ell} Z_k Z_\ell = \sum_{k=1}^n A_{i,k} A_{j,k} Z_k^2 + \sum_{k\neq\ell}^n A_{i,k} A_{j,\ell} Z_k Z_\ell.$$

From the linearity of expectation, and the fact that Z is an independent zero mean unit variance random vector, we get $\Sigma_{i,j} = (AA^T)_{i,j}$. □

Proposition 2.6. *The density for a Gaussian random vector $X : \Omega \to \mathbb{R}^n$ with mean $\mu \in \mathbb{R}^n$ and invertible covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$, is given by*

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right) \text{ for all } x \in \mathbb{R}^n.$$

Proof. We can write $X = \mu + \Sigma^{\frac{1}{2}} Z$, where $Z : \Omega \to \mathbb{R}^n$ is an *i.i.d.* zero mean unit variance Gaussian random vector. Then, we observe that $Z = \Sigma^{-\frac{1}{2}}(X - \mu)$. This implies that the Jacobian matrix $J(z) = \Sigma^{-\frac{1}{2}}$, since the (i,j)th component of the Jacobian matrix $J(z)$ is given by $J_{i,j}(z) = \frac{\partial z_j}{\partial x_i} = \Sigma_{j,i}^{-\frac{1}{2}}, \quad i,j \in [n]$. Recall that the density of Z is $f_Z(z) = \frac{1}{\sqrt{(2\pi)^n}} \exp(-\frac{1}{2} z^T z)$, and from the transformation of random vectors, we get

$$f_X(x) = f_Z(\Sigma^{-\frac{1}{2}}(x-\mu)) \det(\Sigma^{-\frac{1}{2}}) = \frac{1}{(2\pi)^{n/2} \det(\Sigma)^{1/2}} \exp\Big(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\Big), \quad x \in \mathbb{R}^n.$$

□

Remark 11. For any $u \in \mathbb{R}^n$, we compute the characteristic function Φ_X from the distribution of Z as

$$\Phi_X(u) = \mathbb{E}e^{j\langle u, X\rangle} = \mathbb{E}\exp\left(j\langle u,\mu\rangle + j\left\langle A^T u, Z\right\rangle\right) = \exp(j\langle u,\mu\rangle)\Phi_Z(A^T u) = \exp(j\langle u,\mu\rangle - \frac{1}{2}u^T \Sigma u).$$

Lemma 2.7. *If the components of the Gaussian random vector are uncorrelated, then they are independent.*

Proof. If a Gaussian vector is uncorrelated, then the covariance matrix Σ is diagonal. It follows that we can write $f_X(x) = \prod_{i=1}^n f_{X_i}(x_i)$ for all $x \in \mathbb{R}^n$. □