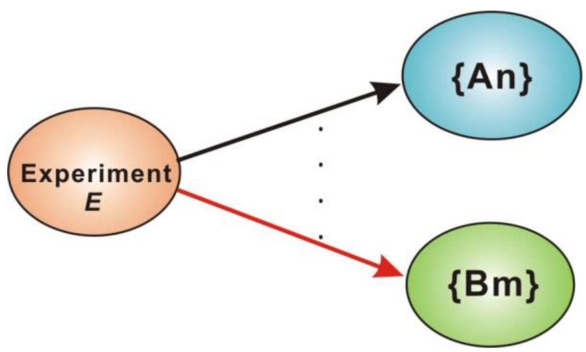
Chapter 2 Introduction to Probability Theory and Random Variables

2.1 Definitions of the Terms

2.1.1 Events and Event Space of an Experiment

We define an experiment $\boldsymbol{E}$ to be a fixed procedure which can be repeated with a directly observable outcome.

Each outcome is arbitrarily associated with an event $\boldsymbol{A}_n$. Thus, the complete set of events $\{\boldsymbol{A}_n\}$ comprises an event space $\boldsymbol{S}$.



For example, $\boldsymbol{E}$=roll a dice,

$\{\boldsymbol{A}_n\}=\{\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_3, \boldsymbol{A}_4, \boldsymbol{A}_5, \boldsymbol{A}_6\}$ or

$\{\boldsymbol{B}_n\}=\{\boldsymbol{B}_1(\text{roll}<3), \boldsymbol{B}_2(\text{roll}=3), \boldsymbol{B}_3(\text{roll}>3)\}$

2.1.2 Definition of Probability

Associated with each possible event A of an experiment $\boldsymbol{E}$ is its probability of occurrence $P(A)$.

Three Axioms of Probability

Axiom 1: $P(A) \geq 0$.

Axiom 2: $P(S) = 1$ for S=a certain event space.

Axiom 3: If A and B are disjoint in an event space, then $P(A \textbf{ or } B) = P(A) + P(B)$.

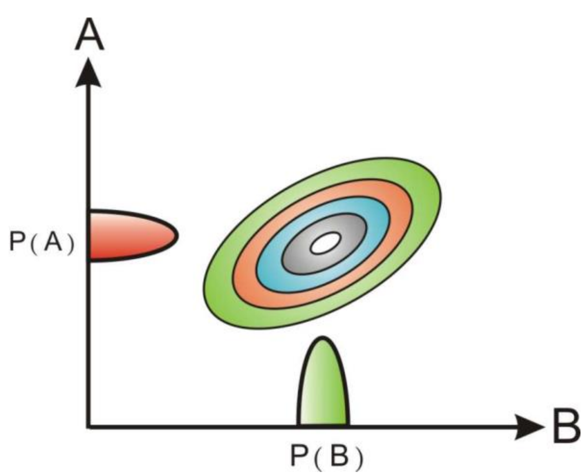
An Intuitive Picture of Probability = frequency of occurrence $f(B)$ of an event B. Consider an experiment is carried through a large number of times $\boldsymbol{N}$, the number of occurrence of an event B is m. Hence

$$f(B) = \textit{frequency of occurence of event } B = (m/N).$$

Thus, based on the **Law of Large Numbers**, we can determine the probability of event B as $f(B) = \textit{probability of the event } B = \lim_{N \to \infty} (m/N)$.

2.1.3 Definitions of Marginal Probability and Conditional Probability

Consider a joint event (denoted as $A_m B_n$) from A_m and B_n. If $\{B_n\}$ be disjoint and form an event space (*i.e.*, $S=\{B_1$ or B_2 or B_3 or ……$B_N\}$), then the event A_m can be viewed as a joint event of A_m and $\{B_1$ or B_2 or B_3 or ……$B_N\}$: (A_m and B_1) or (A_m and B_2) …….or (A_m and B_N) .



We can therefore define the marginal probability of $\boldsymbol{A}_m$ as

$$\therefore P(A_m) = \sum_{n=1}^{N} P(A_m B_n) = \textbf{ marginal probability of } A_m$$

$$\uparrow \textbf{ Joint probability of } (A_m \textbf{ and } B_n)$$

The **conditional probability** of event *B* if event *A* already occurred can be defined as $P(B \mid A) = P(AB) \big/ P(A)$.

However, if knowledge of event *A* has no effect upon the occurrence of *B*, $P(B \mid A) = P(B)$, then events *A* and *B* are called **statistical independent**. That is $P(B \mid A) = P(AB)/P(A) = P(B)$. Therefore, $\boldsymbol{P(AB)=P(A)P(B)}$ (Noted that this is a **necessary condition** for statistical independent).

The following probability laws can be proposed

- **Partition Law:** If $\{B_n\}$ be disjoint and form an event space,

$$P(A_m) = \sum_{n=1}^{N} P(A_m \mid B_n) P(B_n).$$

- **Bayes Rule:**

Since $P(AB) = P(BA)$, thus $P(B \mid A) = \dfrac{P(A \mid B)P(B)}{P(A)}$.

If $B_n \in \{B_m\} =$ **disjoint and form an event space**, then

$$P(B_n \mid A) = \frac{P(A \mid B_n)P(B_n)}{P(A)} = \frac{P(A \mid B_n)P(B_n)}{\sum_m P(A \mid B_m)P(B_m)},$$ indicating that we only require knowledge of quantities $P(B_n)$ and $P(A \mid B_n)$.

See the webpage *for Bayes rule application* in model fitting, https://users.fmrib.ox.ac.uk/~saad/ONBI/bayes_practical.html, or the workshop 2 of this course.

Application of Bayes rule

Assuming the probability of a certain medical test being positive is 90%, if a patient has disease D. A prior knowledge of the disease is that about 1% of the population has the disease, and the test records a false positive 5% of the time. Estimate the probability of having D if a test is positive.

Reformulate the question as follows: $P(+|D)=0.9$, $P(D)=0.01$, $P(+|\text{no } D)=0.05$, calculate $P(D|+)=?$

From Bayes rule, $P(D\,|+)=\dfrac{P(+\,|\,D)P(D)}{P(+)}=\dfrac{P(+\,|\,D)P(D)}{P(+\,|\,D)P(D)+P(+\,|\,no\ D)P(no\ D)}$.

Substituting in the numbers, we obtain $P(D|+) = 0.15$, indicating that the prediction power of the positive test is not so high!

We can apply Bayes rule to create **a machine learning system** that can refine a model ***M*** in the light of the experimental data ***D***, starting from a set of *a priori* knowledge (or assumptions) ***C***. To do that, the first step is to define a conditional prior probability $P(M\,|\,C)$ for a model M (with some model parameters, which are to be refined) based on the initial assumptions from *a priori* knowledge C. Next, we update $P(M\,|\,C)$ in response to the experimental data (***D***) to give the posterior probability $P(M\,|\,D$ and $C)$. The Bayes theorem can be used to offer an estimate of the posterior probability:

$$
\begin{array}{l}
P(M \mid D, C) = \dfrac{P(M \mid C)P(D \mid M, C)}{P(D \mid C)} \\
\text{Prior} : P(M \mid C) \\
\text{Likelihood Probability} : P(D \mid M, C)
\end{array}.
$$

For this rule to be applicable, it must be possible to define the likelihood $P(D \mid M, C)$ that the experimental data ***D*** are consistent with the model ***M*** and the prior assumptions ***C***. The algorithm to implement a Bayesian estimator is typically as follows:

(i). Redefine the previous *a posterior* probability $P_{k-1}(M \mid D_{k-1}\ C)$ as the new *prior* probability via $P_k(M \mid C) = P_{k-1}(M \mid D_{k-1}\ C)$.

(ii). Record the measurement at time *k*, D_k.

(iii). Calculate the *likelihood* $P(D_k \mid M\ C)$ from the model. These could be precomputed, analytically or numerically. This likelihood depends upon the form of the assumed model, which is not necessarily Gaussian.

(iv). For the sake of efficiency, one may want to adjust the numerical range and resolution of *M* considered.

(v). Compute the new (unnormalized) *a posterior* probability via $\tilde{P}(M \mid D_k\ C) = P_k(M \mid C)P(D_k \mid M\ C)$.

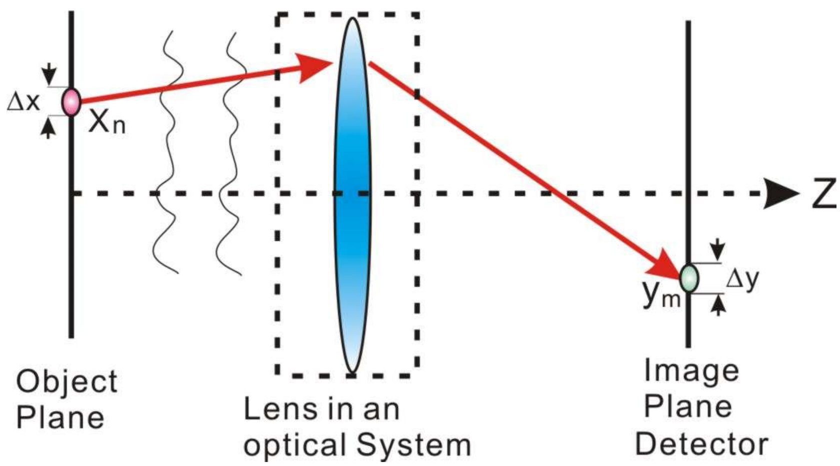
(vi). Normalize to get the new *a posterior* probability $P(M \mid D_k\ C)$.

(vii). Calculate the estimate of the model variables *M* based on the new *a posterior* probability. This choice can be made in several ways, but the simplest approach is to take the maximal value location.

(viii). Repeat at time *k*+1.

This algorithm is the essential core of Bayesian estimation, which becomes very useful since most of the time we do not know what posterior probability is. Bayesian estimation gives a relatively simple way to calculate a posterior probability by multiplying the likelihoods and prior distribution. If we use the point at which the overall *likelihood* is maximal as our estimate, we are performing *maximum likelihood estimation* (MLE). Similarly, we can implement a Maximum a Posteriori (MAP) solver to find a solution, which will maximize the posterior probability.

Application Example of Probability in Optics



Now, let's use an optical imaging system as depicted above to illustrate the concept of probability in optics. First define an object function $o(x_n)$ in the object plane as

$o(x_n) = \dfrac{O(x_n)\Delta x_n}{\sum_n O(x_n)\Delta x_n}$ = photon is emitted from an interval of Δx_n at x_n divided by

total number of emitted photons from the object $= P(x_n)$.

Similarly, the total probability of photons incident on an interval of Δy_m centered at y_m can be described as $P(y_m) = \dfrac{I(y_m)\cdot \Delta y_m}{\sum_n I(y_n)\cdot \Delta y_n} = i(y_m)$.

The point spread function of an optical system can be modelled by a conditional probability $S(y_m; x_n)$, which can be defined as

$S(y_m; x_n) =$ conditional probability $P(y_m \mid x_n)$ that the photon emitted by the object at the position x_n will arrive at y_m on the image plane.

Based on Bayes rule, we obtain $P(x_n \mid y_m) = \dfrac{P(y_m \mid x_n)\cdot o(x_n)}{i(y_m)}$, which can be identified as an inverse point spread function of the system.

2.1.4 Markov Events

If the occurrence probability of an event changes from trial to trial and depends upon the outcome of the preceding trial, the sequence of events is called *Markov events.*

For example, two products ***A*** and ***B*** are competing for sales. Due to the better quality of product A, we found $P(A_{n+1} \mid A_n) = 0.8$ and $P(B_{n+1} \mid B_n) = 0.4$. By using **Axiom 2** of probability $P(A_{n+1} \mid A_n) + P(B_{n+1} \mid A_n) = 1,\ and\ P(A_{n+1} \mid B_n) + P(B_{n+1} \mid B_n) = 1$, we then have

$$
\begin{aligned}
P(A_{n+1} \mid A_n) = 0.8 \quad &\rightarrow \quad P(B_{n+1} \mid A_n) = 0.2 \\
P(B_{n+1} \mid B_n) = 0.4 \quad &\rightarrow \quad P(A_{n+1} \mid B_n) = 0.6
\end{aligned}.
$$

For $n \rightarrow \infty, \quad P(A) = ?$

To solve the question, we first apply the Partition Law of Probability:

$$P(A_{n+1}) = P(A_{n+1} \mid A_n)P(A_n) + P(A_{n+1} \mid B_n)P(B_n)$$
$$= 0.8\, P(A_n) + 0.6\, P(B_n)$$
$$P(B_{n+1}) = P(B_{n+1} \mid A_n)P(A_n) + P(B_{n+1} \mid B_n)P(B_n)$$
$$= 0.2\, P(A_n) + 0.4\, P(B_n)$$

As $n \to \infty$

$P(A) = 0.8\, P(A) + 0.6\, P(B)$, $P(B) = 0.2\, P(A) + 0.4\, P(B)$ and

$P(A) + P(B) = 1$.

Therefore, we determine $P(A) = 0.75, \quad P(B) = 0.25$.

2.2 Continuous Random Variables

2.2.1 Definition of a Random Variable (RV)

Definition of a random variable ***U***: By examining the outcome of an experiment ***E***, we assign a number (real or complex) $u(A_n)$ to every possible elementary event A_n. Thus, a random variable ***U*** consists of all possible **$\{u(A_n)\}$** together with an associated measure of their probabilities **$P(A_n)$**.

Probability distribution function $F_U(u)$ of a RV ***U***: $F_U(u)$=Probability$\{U \le u\}$.

Axiom 1: $P(A) \ge 0 \quad \Rightarrow \quad F_U(u)$ is nondecreasing to the right.

Axiom 2: $P(C) = 1 \quad \Rightarrow \quad F_U(+\infty) = 1$ and $P(N) = 0 \quad \Rightarrow \quad F_U(-\infty) = 0$.

Therefore,

$$P_U(u) = probability\ density\ \ function, pdf\ = \frac{\partial F_U(u)}{\partial u} = \lim_{\Delta u \to 0} \frac{F_U(u) - F_U(u-\Delta u)}{\Delta u}.$$

For a discrete RV $\{u_k\}=\{u(A_k)\}$, we can express the associated *pdf* as

$$P_U(u) = \sum_{k=1}^{\infty} P(u_k)\delta(u-u_k).$$

2.2.2 Statistical Average and Moments of a Random Variable

If ***U*** is a RV, $\xrightarrow[\substack{function\\ mapping,g}]{}$ $g(\boldsymbol{U})$ is also a RV.

Note:

$$\bar{g}(u) = E[g(u)] = \int_{-\infty}^{+\infty} g(u)P_U(u)du = \int_{-\infty}^{+\infty} g(u)\sum_n P(u_n)\delta(u-u_n)du$$

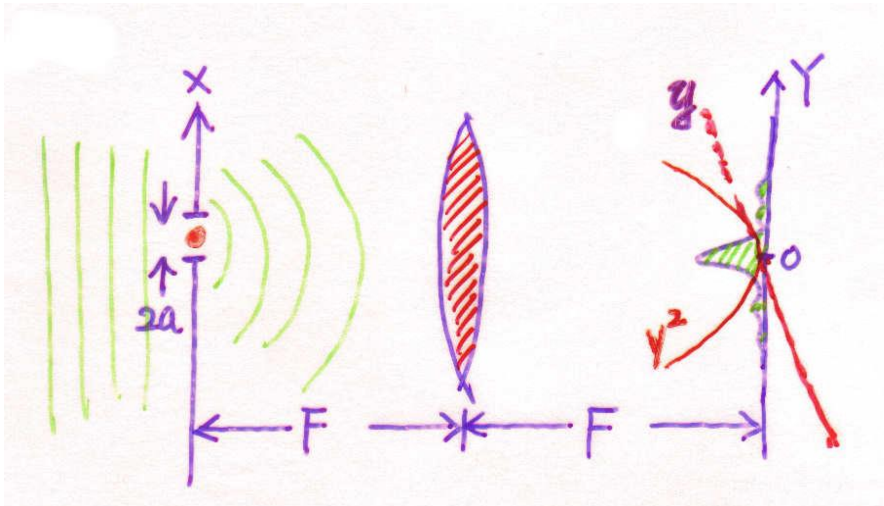
$$\overset{discrete}{=} \sum_n g(u(A_n))P_U[u(A_n)] = statistical\ average\ of\ \ g(u).$$

- If $g(u) = u^n$, $\overline{u^n} = \int_{-\infty}^{+\infty} u^n P_U(u)du =$ n-th moment
- If $g(u) = (u-\bar{u})^n$, $E[(u-\bar{u})^n] = \overline{(u-\bar{u})^n} =$ *n*-th central moment. When *n*=2, $\sigma^2 = \overline{(u-\bar{u})^2} =$ variance=2nd central moment with σ=standard deviation.
- If $g(u_1, u_2, ..., u_N) = \sum_{i=1}^{N} a_i u_i$, then

  $$\bar{g} = E[g] = \sum_{i=1}^{N} a_i \bar{u}_i = \int ... \int \sum_i a_i u_i P(u_1, ..., u_N) du_1 ... du_N$$

Application Example of Probability in Optics

Consider an imaging system,



$$i(y_m) = \sum_n S(y_m; x_n) o(x_n) \quad \underset{\substack{\Delta x \to 0 \\ \Delta y \to 0}}{\longrightarrow} \quad i(y) = \int_{-\infty}^{+\infty} dx\, s(y; x) o(x)$$

The point spread function (PSF), $S(y) = \frac{ka}{\pi F} \mathrm{sinc}^2(\frac{ka}{F} y)$, is in fact a probability density function for position y of a photon in the image plane if it originates at a point located at the origin in the object plane.

The moments of a random variable are

$$m_1 = \bar{y} = \int_{-\infty}^{+\infty} y\, S(y) dy = \int_{-\infty}^{+\infty} y\, \frac{ka}{\pi F} \mathrm{sinc}^2(\frac{ka}{F} y) dy = 0$$

$$m_2 = \overline{y^2} = \int_{-\infty}^{+\infty} y^2\, S(y) dy = \infty .$$

(i) $i(y_m)=\sum_{n=1}^{N} S(y_m;x_n)o(x_n)$: If $o(x_n)$ is a RV (e.g., a randomly selected member of a set of objects), then $i(y_m)$ is also a RV and

$$E[i(y_m)]=\sum_{n=1}^{N} S(y_m;x_n)E[o(x_n)].$$

(ii) If $S(y_m;x_n)$ is a RV (e.g., an object is imaged repeatedly through atmospheric turbulences, a speckle pattern will be generated from $S(y_m;x_n)$), then $E[i(y_m)]=\sum_{n=1}^{N} E[S(y_m;x_n)]o(x_n)$.

2.2.3 Useful Probability Laws in Optics

A. Poisson Distribution

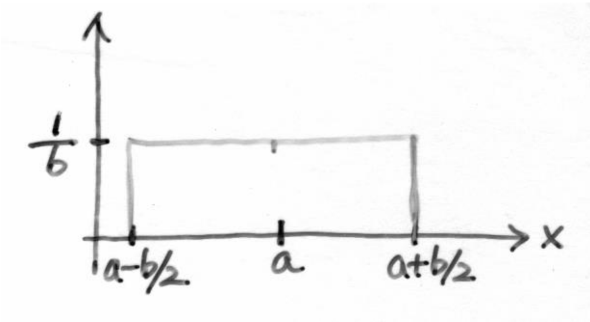
$$P(A_n)=p_n=\frac{a^n}{n!}e^{-a} \quad n=0,1,2,....\ with\ \ a>0.$$

Here a is the sole parameter of Poisson distribution, which determines the mean, variance, and even the third central moment.

This probability law arises for n photon arrivals over a time interval T if the photons (1) arrive uniformly (with an average arrival rate of a/T) and randomly in time; (2) arrive rarely; and (3) arrive independently.

B. Binomial

The binomial law arises under the same circumstances that produce the Poisson law. However, it does not require rare events, as does the Poisson.



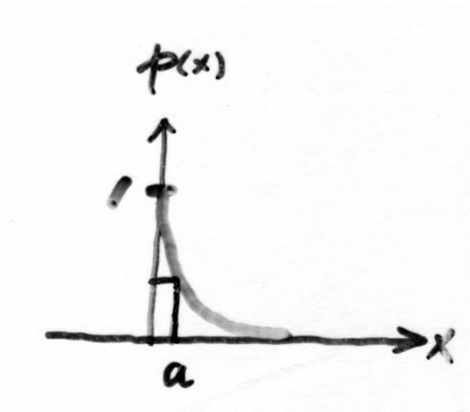
$$P(A_n) = p_n = \binom{N}{n} \alpha^n \beta^{N-n}, \quad \alpha + \beta = 1$$

The average and central moment can be calculated as

$$\bar{n} = \sum_{n=0}^{\infty} n\, p_n = N\alpha \ , \ \ \sigma^2 = N\alpha\beta \, .$$

C. Uniform

$$P(x) = \frac{1}{b} \mathrm{Re}\, ct(\frac{x-a}{b}) \, .$$



The average and 2nd central moment can be calculated to be

$$\bar{x} = a, \quad \sigma^2 = b^2/12 \, .$$

This probability is useful to depict the photon statistical law at position x on a uniformly bright object.

D. Exponential

$$P(x) = \begin{cases} \frac{1}{a} e^{-x/a} & x \geq 0 \\ 0 & x < 0 \end{cases} .$$

The average and 2nd central moment can be calculated to be

$$\bar{x} = a, \quad \sigma^2 = a^2 .$$

This law arises as the probability density for an intensity (a random variable) $x=I$ in laser speckle, with a mean intensity $\overline{I}=a$.

E. 1-D Normal Distribution

$$P(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\overline{x})^2/(2\sigma^2)}.$$

This is a very useful probability distribution due to the central limit theorem holds for most of the physical processes. For example, the optical phase after passing through atmospheric turbulance obeys a normal law.

F. 2-D Normal Distribution

$$P(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\overline{x})^2}{\sigma_1^2}+\frac{(y-\overline{y})^2}{\sigma_2^2}-\frac{2\rho(x-\overline{x})(y-\overline{y})}{\sigma_1\sigma_2}\right]}$$, where

$$\sigma_1^2=\overline{x^2}-(\overline{x})^2$$

$$\sigma_2^2=\overline{y^2}-(\overline{y})^2$$

$$\rho=\overline{(x-\overline{x})(y-\overline{y})}\Big/(\sigma_1\sigma_2)=\textit{cross correlation between RVs } x \textit{ and } y$$

2.4 Fouriers Methods

This section aims to introduce the Fourier analysis to RVs and probability theory. The basic concept is easy to understand by referring to the following analog

Linear Optics ↔ **Statistical Optics** : Statistical nature of Optics

(**Fourier Optics**) ↔ Fourier analysis on RV U and probability theory

Optical Signal $E(k)$ ↔ *pdf* $P_U(u)$

2.4.1 Characteristic Function

$$\varphi_U(\omega) = \int P_U(u) \cdot e^{i\omega u} du = \text{characteristic function for RV } \boldsymbol{U}.$$

↑

pdf of a RV U at the value u

$\omega \leftrightarrow u$ forms a conjugate pair.

Note: $\varphi_U(\omega) = E[e^{i\omega u}]$ with E[..]=statistical averaging.

2.4.2 Applications of Characteristic Function

A. Generating Moments

Note: $$\frac{\partial^n \varphi_U(\omega)}{\partial \omega^n}\Big|_{\omega=0} = \frac{\partial^n}{\partial \omega^n} E[e^{j\omega u}]\Big|_{\omega=0} = E[(ju)^n] = j^n E[u^n] = j^n m_n .$$

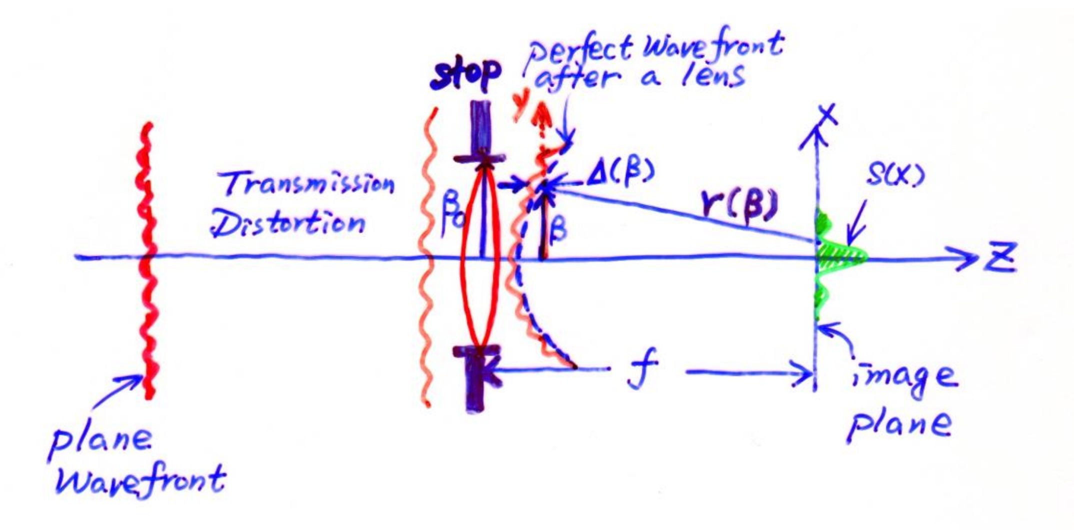
i.e., the behavior of $\varphi_U(\omega)$ at the origin (ω =0) defines all the moments of $P_U(u)$.

B. Describing RVs

Note: $RV = \{u \ and \ P_U(u)\}$.

$P_U(u) = \frac{1}{2\pi}\int \varphi_U(\omega)\cdot e^{-j\omega u} d\omega$, which is the inverse FT of characteristic function. Thus, characteristic function can fully determine $P_U(u)$ and therefore the RV.

Application Example of Probability in Optics



The reduced coordinate β on the pupil plane $\beta = \frac{2\pi}{\lambda}\frac{y}{f} = ky/f$ denoting the normal component of k when y is the transverse coordinate on the lens.

$s(x)$ = point spread function = $|a(x)|^2$, where $a(x)$=Point amplitude function.

From Huygens' principle: $a(x) = \int_{-\beta_0}^{+\beta_0} \frac{e^{jkr(\beta)}}{r(\beta)} d\beta$.

By noting that

$r^2 = (y-x)^2 + [f+\Delta(\beta)]^2 \quad and \; |f+\Delta(\beta)| \gg |(y-x)|$,

we can obtain $a(x) \approx \int_{-\beta_0}^{+\beta_0} e^{jk\Delta(\beta)+j\beta x} d\beta$.

The optical transfer function (**OTF**) $T(\omega)$ of the system can also be deduced to be

$$T(\omega)=\int_{-\infty}^{+\infty} s(x)e^{-j\omega x}dx \Big/ \int_{-\infty}^{+\infty} s(x)dx = \frac{1}{2\beta_0}\int_{\omega-\beta_0}^{\beta_0}[e^{jk\Delta(\beta)} * e^{jk\Delta(\omega-\beta)}]d\beta$$

= convolution of the pupil function $e^{jk\Delta(\beta)}$.

Let $\omega \leftrightarrow k \quad u \leftrightarrow \Delta(\beta)$, we find $\varphi_U(\omega)=E[e^{j\omega u}] \rightarrow E[e^{jk\Delta(\beta)}]$, denoting that $\varphi_U(\omega)$ is an average amplitude of the field with a random fluctuating phase front.

$T(\omega)=FT[s(x)]$ and $s(x)$= probability law of photon position on the image plane. Thus, $T(\omega)$ can be understood as the characteristic function for probability law $s(x)$.

2.4.3 Shift Theorem

For a random variable ***U***, we can generate a RV ***W*** as

$$W = aU + b.$$

Therefore, we deduce the characteristic function of ***W*** as

$$\begin{aligned}\Phi_{W=aU+b}(w) &= E[e^{jwu}] = E[e^{jw(au+b)}] = e^{jwb}E[e^{j(wa)u}] \\ &= e^{jwb}\Phi_U(aw)\end{aligned}.$$

2.4.4 Characteristic Functions for Some Probability Laws

(a) Poisson

Let $p_U(u) = \sum_{n=0} p_n \delta(u-n)$ *and* $p_n = \frac{a^n}{n!} e^{-a}$ *for* $n = 0,1,2,....$ *and* $a > 0$.

Then the corresponding characteristic function becomes

$$\Phi_U(w) = \int_{-\infty}^{+\infty} e^{jwu} p_U(u) du = e^{-a} \sum_{n=0}^{\infty} [\frac{a^n}{n!} e^{jwn}] = e^{-a} \cdot e^{(ae^{jw})}.$$

We can then use $\Phi_U(w)$ to obtain the successive moments of the Poisson RV ***U***.

(b) Binomial Law

Similarly by use of

$$\Phi_U(w) = \sum_{n=0}^{N} e^{jwn} \binom{N}{n} \alpha^n \beta^{N-n} = (\alpha e^{jw} + \beta)^N$$, where $\alpha + \beta = 1$.

By use of the characteristic function, we obtain

$$m_1 = first\ moment = -j \frac{\partial \Phi_U(w)}{\partial w}\Big|_{w=0} = -jN(\alpha e^{jw} + \beta)^{N-1}(j\alpha e^{jw})\Big|_{w=0} = N\alpha$$

and

$$\sigma_m^2 = 2nd\ central\ moment = N\alpha\beta.$$

(c) Uniform Case

Here we just list the result for the characteristic function for the uniform RV with

$$P(x) = \frac{1}{b} \mathrm{Re}ct(\frac{x-a}{b}),$$

$$\Phi_U(w) = e^{jwa} \cdot \text{sinc}(\frac{bw}{2}).$$

(d) Exponential Case

The characteristic function for the exponential law $P(x) = \begin{cases} \frac{1}{a}\ e^{-x/a} & x \geq 0 \\ 0 & x < 0 \end{cases}$

is $\Phi_U(w) = \dfrac{1}{1 - jwa}$.

(e) 1D Normal Distribution Case

The characteristic function becomes

$$\Phi_U(w) = e^{jw\bar{u} - \sigma_u{}^2 w^2 / 2} \quad \textit{where}$$

$\bar{u} = \textit{first moment}$ and $\sigma_u^2 = \textit{2nd central moment}$.

(f) 2D Normal Distribution Case (two correlated, bivariate RU)

Let $\vec{U} = \{U_1, U_2, \ldots, U_N\}$ be a N-dimensional RV, then

the corresponding characteristic function becomes

$$\Phi_{\vec{U}}(\vec{\omega}) = \Phi_{\vec{U}}(\omega_1, \omega_2, \ldots, \omega_N) = \int d\vec{u}\ e^{j\vec{\omega}\cdot\vec{u}} p_{\vec{U}}(\vec{u}).$$

For 2D normal case, $\vec{U} = \{U_1, U_2\}$

$$\Phi_{\vec{U}}(\omega_1,\omega_2) = \int d\vec{u}\ e^{j\vec{\omega}\cdot\vec{u}} p_{\vec{U}}(\vec{u})$$
$$= e^{j\omega_1\bar{u}_1 + j\omega_2\bar{u}_2 - \frac{1}{2}(\sigma_1{}^2\omega_1{}^2 + \sigma_2{}^2\omega_2{}^2 + 2\rho\,\sigma_1\sigma_2\omega_1\omega_2)}$$

$$\rho = correlation\ coefficient\ between\ U_1\ and\ U_2\,.$$

We will apply this result to find the long-term average optical transfer function due to turbulence.

2.4.5 Probability Law for the Sum of Two Independent Random Variables

Let ***U*** and ***V*** be two statistically independent RVs. If ***W***=***U***+***V***, ***W*** will be a RV too.

But what is the corresponding pdf, $p_W(w)$?

Note that

$$\Phi_W(\omega) = E[e^{j\omega w}] = E[e^{j\omega(u+v)}] \xrightarrow{S.I.} E[e^{j\omega u}]E[e^{j\omega v}] = \Phi_U(\omega)\,\Phi_V(\omega)\,.$$

Take an inverse Fourier transform of $\Phi_W(\omega)$,

$$p_W(w) = \frac{1}{2\pi}\int \Phi_W(\omega)\ e^{-j\omega w} d\omega = \int \Phi_U(\omega)\ \Phi_V(\omega)\ e^{-j\omega w} d\omega$$
$$= p_U(w) \otimes p_V(w)\,.$$

If ***U*** and ***V*** are two statistically independent ($i.e.,\ \rho = 0$) Gaussian RVs, then ***W*** is a Gaussian RV too and

$\Phi_W(\omega) = e^{-(\sigma_u{}^2 + \sigma_v{}^2)\omega^2/2}$, *i.e.*, $p_W(w)$ will be broader than p_U and p_V.

Now let us consider an image formation system, where

$o(x)$ = probability of photons emitted at *x* in the object plane, and

$i(y)$ = probability of photons arrived at *y* in the image plane.

Let *x*, *y* be RVs, and *y*=*x*+(*y*-*x*), so (*y*-*x*) is also a RV, which denotes an incremental displacement to the side with a probability density law of *s*(*y*; *x*).

If (*y*-*x*) is statistically independent of *x*, which is valid when object is small, then

$s(y;x) = s(y-x)$. This condition is called *isoplanatism* in image forming theory, or is called *strict-sense stationary* in statistical theory.

Thus, $y = x + (y - x)$ is a RV $\rightarrow$ $i(y) = s(x) \otimes o(x)$ is the associated probability law.

It is interesting to ask:

Is it possible to make $s(x)$ *negative going such that* $i(y)$ *is narrower than* $o(x)$?

The answer is *Yes*, which is essentially an image enhancement or restoration procedure.

What are the resulting mean and variance of the sum of *N* statistically independent RVs ?

Assume $\{U_m\}$ in $W = U_1 + U_2 + ... + U_N$ to be statistically independent and normal.

Thus, $\Phi_W(w) = \Phi_{U_1}(w)\Phi_{U_2}(w)...\Phi_{U_N}(w)\ with\ \ \Phi_{U_m}(\omega) = e^{j\omega\bar{u}_m - \frac{1}{2}\sigma_m{}^2\omega^2}$.

We obtain

$$\Phi_W(\omega) = e^{j\omega \sum_{m=1}^{N} \bar{u}_m - \frac{1}{2}\sum_{m=1}^{N} \sigma_m{}^2 \omega^2}$$

, indicating that ***W*** is also a normal RV with

$$\bar{W} = \sum_{m=1}^{N} \bar{u}_m \text{ and } \sigma_W^2 = \sum_{m=1}^{N} \sigma_m{}^2 .$$

In general, $\{U_m\}$ can be **statistically independent** and follows any distribution,

$\bar{W} = \sum_{m=1}^{N} \bar{u}_m$ is a normal RV and $\sigma_W^2 = \sum_{m=1}^{N} \sigma_m{}^2$.

2.4.6 Central Limit Theorem

Let us assume $\{U_m\}$ to be

(1) *n* statistically independent RVs, obeying

(2) identical characteristic function $\Phi_U(w)$.

We first define $W = U_1 + U_2 + ... + U_n$, and will prove $p_W(w)$ to become Gaussian when $n \to \infty$.

For simplicity, let us assume the means of all $\{U_m\}$ to be zero. Then, we have

$$\begin{aligned}\Phi_W(w) &= [\Phi_U(w)]^n = [\Phi_U(0) + w\Phi_U{}'(0) + \frac{w^2}{2}\Phi_U{}''(0) + \mu w^3]^n \\ &= [1 - \frac{\sigma_u{}^2 w^2}{2} + \mu w^3]^n\end{aligned}.$$

Defining a new RV ***S*** as

$S = \frac{W}{\sqrt{n}} = \frac{U_1}{\sqrt{n}} + \frac{U_2}{\sqrt{n}} + \ldots + \frac{U_n}{\sqrt{n}}$ and by using the Shift theorem, we find the corresponding characteristic function of ***S*** as

$$\Phi_S(w) = \Phi_W(w/\sqrt{n}) = [1 - \frac{\sigma_u{}^2 w^2}{2n} + \frac{\mu w^3}{n^{3/2}}]^n .$$

Let us calculate the logarithmic function of $\Phi_S(w)$,

$$\ln \Phi_S(w) = n \ln(1 - \frac{\sigma_u{}^2 w^2}{2n} + \frac{\mu w^3}{n^{3/2}}].$$

When $n \to \infty$, $\ln \Phi_S(w) = n \ln(1 - \frac{\sigma_u{}^2 w^2}{2n} + \frac{\mu w^3}{n^{3/2}} + O(\frac{\sigma_u{}^4 w^4}{n^2}] \to -\frac{\sigma_u{}^2 w^2}{2}$.

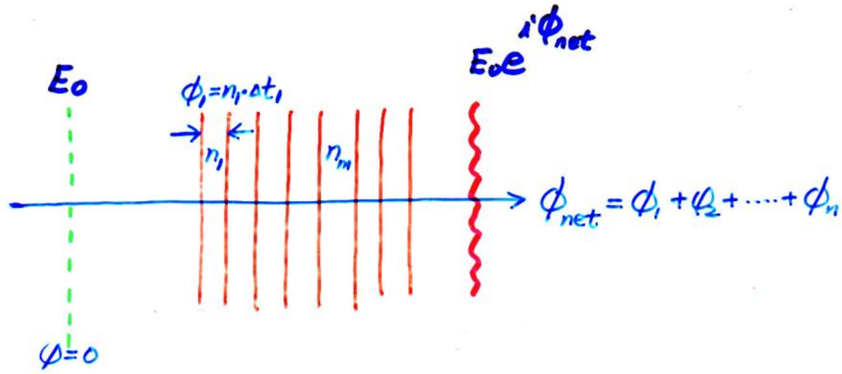
$$\therefore \Phi_S(w) \overset{n\to\infty}{\to} e^{-\frac{\sigma_u{}^2 w^2}{2}} .$$

In fact, $\{U_m\}$ do not have to all obey the same probability law and they do not all have to be independent.

Optical Application of the Central Limit Theorem

We can apply the central limit theorem on modeling the effect of atmospheric turbulence.

The light wave from a distant star behaves like a planar wavefront near the Earth. Let us first divide the atmosphere of the Earth along the optical path into N slabs with an index of refraction n_m and thickness Δt_m in the *m*th slab. The phase delay of the wavefront after reaching the ground can be approximated by

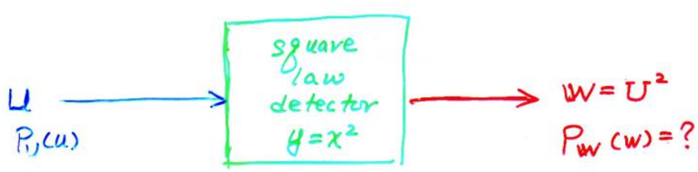


$\varphi_m = \frac{2\pi}{\lambda} n_m \cdot \Delta t_m$ with n_m denoting the random index of refraction of the mth plane and $c \cdot \Delta t$ being the thickness. We note that $\{\varphi_m\}$ form a set of RVs with

$$\varphi_{net} = \sum_{m=1}^{N} \varphi_m \underset{N \to \infty}{\longrightarrow} Gaussian\ RV .$$

Thus, $p(\varphi_{net})$ is Gaussian probability, and $p(\varphi_{net}(t), \varphi_{net}(t'))$ a bivariate Gaussian.

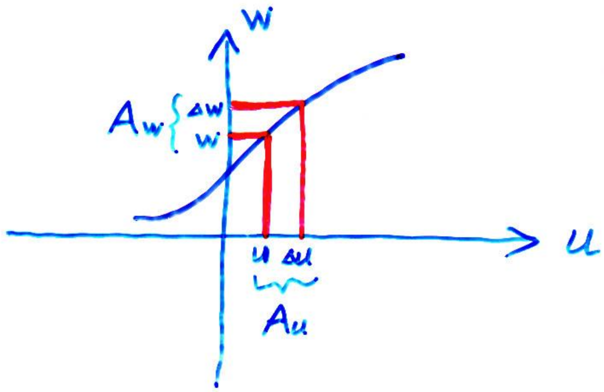
2.5 Functions of Random Variables



2.5.1 Single RV

Let $W = f(U)$ and U be a RV with a known *pdf* $p_U(u)$. The inverse function $u = f^{-1}(w)$ can possess either (A) a unique root u_1 or (B) a multiplicity of roots u_1, u_2,..., u_r.

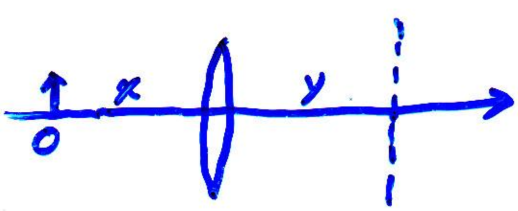
(A)We first consider the case with unique root.



Now convert the event space of $A_w = \{w \leq W \leq w + dw\}$ into $A_u = \{u \leq U \leq u + du\}$. For each event in A_w, it may be alternatively described as a corresponding event in A_u, *i.e.*, $p(A_w) = p(A_u)$. The relative number of times a value *w* will occur equals the relative number of times the corresponding value of *u* will occur. Therefore,

$$p_W(w)dw = p_U(u)du = p_U[f^{-1}(w)]\frac{du}{dw}dw \quad \Rightarrow p_W(w) = \frac{p_U[u = f^{-1}(w)]}{\left|\frac{dw}{du}\right|}.$$

Application Example in Optics



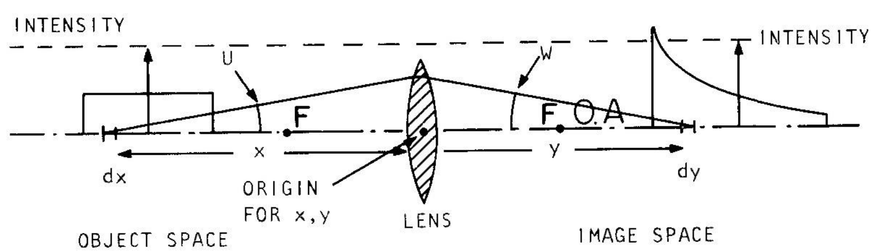
Consider an imaging system with a simple lens $\frac{1}{y} + \frac{1}{x} = \frac{1}{f}$.

If $p_X(x)$ is known for an object, then

$x = f^{-1}(y) = \frac{f\,y}{y-f}$ and $\frac{dy}{dx} = -\frac{(y-f)^2}{f^2}$, leading to $p_Y(y) = \frac{f^{\,2}}{(y-f)^2}\,p_X(\frac{fy}{y-f})$.

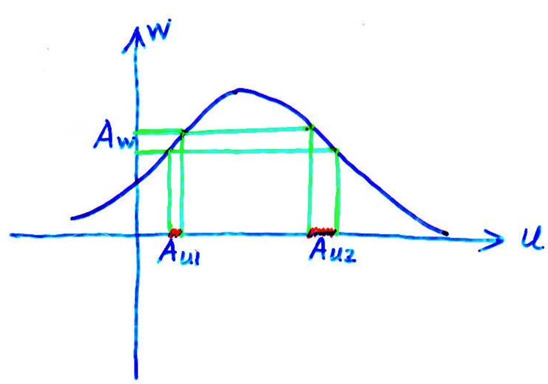
If $p_X(x) = \begin{cases} \frac{1}{f}, & \frac{3f}{2} \le x \le \frac{5f}{2} \\ 0, & \textit{for all other } x \end{cases}$, therefore

$$p_Y(y) = \begin{cases} \frac{f}{(y-f)^2}, & \frac{5f}{2} \le y \le 3f \\ 0, & \textit{elsewhere} \end{cases}.$$



(B) Next, consider the case with multiple roots

For example, for optical detection with a square-law detector $W = U^2$, we have $U = \pm\sqrt{W}$. Therefore, $p(A_W) = p(A_{U_1}) + p(A_{U_2})$, which yields



$$p_W(w) = \frac{p_U(u_1)}{\left|\frac{dw}{du}\right|_{u_1}} + \frac{p_U(u_2)}{\left|\frac{dw}{du}\right|_{u_2}}.$$

Here $u_1 = -\sqrt{w}$ *and* $u_2 = \sqrt{w}$

$$\left.\frac{dw}{du}\right|_{u_1} = 2u_1 = -2\sqrt{w} \quad and$$
$$\left.\frac{dw}{du}\right|_{u_2} = 2u_2 = 2\sqrt{w} \qquad .$$

Thus, we have $p_W(w) = \frac{p_U(-\sqrt{w})}{-2\sqrt{w}} + \frac{p_U(\sqrt{w})}{2\sqrt{w}}$.

If $p_U(u) = \frac{1}{\sqrt{2\pi}\,\sigma_u} e^{-\frac{u^2}{2\sigma_u{}^2}}$, then $p_W(w) = \begin{cases} \frac{1}{\sqrt{w}\sqrt{2\pi}\,\sigma_u} e^{-\frac{w}{2\sigma_u{}^2}}, & w \geq 0 \\ 0, & w < 0 \end{cases}$.

Now let us consider N random variables $\vec{U} = \{U_1, U_2, ..., U_N\}$ with a known joint probability $p_{\vec{U}}(u_1, u_2, ..., u_N)$. We first perform a functional mapping with $\vec{W} = \{W_1, W_2, ..., W_N\} = f(\vec{U})$ and want to know what is the *pdf* of $\vec{W}$.

To find the answer, let us assume $\vec{U} = f^{-1}(\vec{W})$ to possess ***r*** roots for each RV $U_m : u_{m1}, u_{m2}, ..., u_{mr}$.

For simplicity, let ***N***=2, ***r***=2, by following the general transformation rule: $p_{\vec{W}}(w_1, w_2, ..., w_N) = |J|\, p_{\vec{U}}[u_1(\vec{W}), u_2(\vec{W}), ..., u_N(\vec{W})]$, we obtain

$$p_{\vec{W}}(w_1, w_2) = p_{\vec{U}}(u_{11}, u_{21})du_{11}du_{21} + p_{\vec{U}}(u_{11}, u_{22})du_{11}du_{22}$$
$$+p_{\vec{U}}(u_{12}, u_{21})du_{12}du_{21} + +p_{\vec{U}}(u_{12}, u_{22})du_{12}du_{22} .$$

Application Example in Optics

Consider a light amplitude ***u*** which is composed of ***N*** random phasors

$$u = a\,e^{j\theta} = R + jN = \frac{1}{\sqrt{N}}\sum_{k=1}^{N}\alpha_k\, e^{j\varphi_k} .$$

To deduce useful conclusions, we make some assumptions:

(1) $\alpha_k \big/ \sqrt{N}$ and φ_k are statistically independent;

(2) The RV α_k is identically distributed for all ***k*** with mean $\bar{\alpha}$ and 2nd moment $\overline{\alpha^2}$;

(3) The RV φ_k is identically distributed in $\left[-\pi, \pi\right]$.

Based on these assumptions, and

$$R = \frac{1}{\sqrt{N}}\sum_{k=1}^{N}\alpha_k \cos\varphi_k \ , \ I = \frac{1}{\sqrt{N}}\sum_{k=1}^{N}\alpha_k \sin\varphi_k \ ,$$

we have

$$\bar{R} = \frac{1}{\sqrt{N}}\sum_{k=1}^{N}\overline{\alpha_k \cos\varphi_k} \overset{SI}{=} \frac{1}{\sqrt{N}}\sum_{k=1}^{N}\overline{\alpha_k}\cdot\overline{\cos\varphi_k} = \sqrt{N}\cdot\bar{\alpha}\cdot\overline{\cos\varphi} = 0$$

$$\bar{I} = \frac{1}{\sqrt{N}}\sum_{k=1}^{N}\overline{\alpha_k \sin\varphi_k} \overset{SI}{=} \frac{1}{\sqrt{N}}\sum_{k=1}^{N}\overline{\alpha_k}\cdot\overline{\sin\varphi_k} = \sqrt{N}\cdot\bar{\alpha}\cdot\overline{\sin\varphi} = 0$$

and

$$\overline{R^2} = \frac{1}{N}\sum_{k=1}^{N}\ \sum_{n=1}^{N}\overline{\alpha_k\alpha_n}\cdot\overline{\cos\varphi_k\,\cos\varphi_n} = \frac{1}{N}\sum_{k=1}^{N}\ \sum_{n=1}^{N}\overline{\alpha_k\alpha_n}\cdot\frac{1}{2}\delta_{kn} = \frac{\overline{\alpha^2}}{2} = \sigma^2$$

$$\overline{I^2} = \frac{1}{N}\sum_{k=1}^{N}\ \sum_{n=1}^{N}\overline{\alpha_k\alpha_n}\cdot\overline{\sin\varphi_k\,\sin\varphi_n} = \frac{1}{N}\sum_{k=1}^{N}\ \sum_{n=1}^{N}\overline{\alpha_k\alpha_n}\cdot\frac{1}{2}\delta_{kn} = \frac{\overline{\alpha^2}}{2} = \sigma^2 .$$

Because correlation between ***r*** and ***i*** is $\rho = 0$, we obtain $\overline{RI} = 0$.

According to the central limit theorem, we have $p_{ri}(R,I) \overset{N\to\infty}{\to} \dfrac{1}{2\pi\sigma^2} e^{-(R^2+I^2)/(2\sigma^2)}$.

Let $a = \sqrt{R^2 + I^2}, \theta = \tan^{-1}(I/R)$ (*i.e.*, $R = a\cos\theta \quad I = a\sin\theta$)

$$\left|J\right| = \begin{vmatrix} \dfrac{\partial R}{\partial a} & \dfrac{\partial R}{\partial \theta} \\ \dfrac{\partial I}{\partial a} & \dfrac{\partial I}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -a\sin\theta \\ \sin\theta & a\cos\theta \end{vmatrix} = a.$$

The joint probability becomes

$$p_{\mathrm{A}\Theta}(a,\theta) = p_{ri}(R = a\cos\theta, I = a\sin\theta)\cdot a = \begin{cases} \dfrac{a}{2\pi\sigma^2}\cdot e^{-a^2/(2\sigma^2)}, & -\pi < \theta \le \pi \\ 0, & otherwise \end{cases}.$$

Form the joint probability density function, we can deduce the marginal probability of ***A*** to be

$$p_{\mathrm{A}}(a) = \int_{-\pi}^{\pi} p_{\mathrm{A}\Theta}(a,\theta)\, d\theta = \begin{cases} \dfrac{a}{\sigma^2}\cdot e^{-a^2/(2\sigma^2)}, & a > 0 \\ 0, & otherwise \end{cases},$$

which is also noted to be **Rayleigh density function**. Similarly, for the marginal probability of Θ, we have

$$p_{\Theta}(\theta) = \int_{0}^{\infty} p_{\mathrm{A}\Theta}(a,\theta)\, da = \begin{cases} \dfrac{1}{2\pi}, & -\pi < \theta \le \pi \\ 0, & otherwise \end{cases}.$$