## Handout 1

Review:

- Distribution function, $F_X(x)$, density function $f_X(x)$, probability function $P_X(x)$.
- Expectation $E[g(X)]$, moments.
- Taylor expansion:

$$g(x) = g(a) + g'(a)(x-a) + \frac{1}{2!}g''(a)(x-a)^2 + \frac{1}{3!}g'''(a)(x-a)^3 + \ldots$$

- When $a = 0$ we have the Maclaurin expansion.
- A particular case we need:

$$e^t = 1 + t + \frac{1}{2!}t^2 + \frac{1}{3!}t^3 + \frac{1}{4!}t^4 + \ldots$$

- Characteristic,

$$\psi_X(u) = E[e^{iuX}]$$

  and moment-generating,

$$\varphi_X(u) = E[e^{uX}]$$

  functions. Relationship to moments:

$$\begin{aligned} \varphi_X(u) &= E[1 + uX + \frac{1}{2!}u^2X^2 + \frac{1}{3!}u^3X^3 + \ldots] \\ &= 1 + \alpha_1 u + \frac{1}{2!}\alpha_2 u^2 + \frac{1}{3!}\alpha_3 u^3 + \ldots \end{aligned}$$

- Quite clearly,

$$\varphi_{aX+b}(u) = E[e^{aX+b}] = e^b \times E[e^{aX}] = e^b \varphi_X(ua)$$

  and similarly for $\psi_{aX+b}$.
- Reminder convergence in probability, convergence in distribution ("in law") and the Central Limit Theorem (CLT).
- Newton's binomial formula:

$$(a+b)^n = \binom{n}{0}a^n b^0 + \binom{n}{1}a^{n-1}b^1 + \binom{n}{2}a^{n-2}b^2 + \ldots + \binom{n}{n}a^0 b^n$$

  (Remember: $\binom{n}{x} = \frac{n!}{x!(n-x)!}$.)

Binary distribution, $b(p)$:

- Simplest possible distribution for a two-state random variable.

- Probability function:

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } q = 1 - p \end{cases}$$

- 0 and 1, conventional labels: could be heads/tails, success/failure, life/death, yes/no, etc.
- Quite easily $E[X] = p$, $\mathrm{Var}(X) = pq$. (What is the largest possible variance?)
- The characteristic function is $\psi(u) = E[e^{iuX}] = q + pe^{iu}$ (if you prefer the moment generating function, it would be $\varphi(u) = E[e^{uX}] = q + pe^{u}$).

Binomial distribution, $b(p, n)$:

- $X \sim B(p, n)$ if $X = X_1 + \ldots + X_n$ and $X_i$ i.i.d as $b(p)$.
- Examples of common situations which can be described in terms of the binomial distribution.
- What is the probability function? Two ways:
  - Simple combinatorial argument: $P_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$.
  - Use the characteristic (or moment-generating) function:

    $$\varphi_X(u) = \varphi_{X_1}(u) \times \cdots \times \varphi_{X_n}(u) = (q + pe^u)^n$$

    We have:

    $$\varphi_X(u) = P_X(0)e^0 + P_X(1)e^u + \ldots + P_X(n)e^{nu} = (q + pe^u)^n$$

    and equating coefficients after applying Newton's binomial formula to the right hand side gives again the previous answer.
- If $X$ and $Y$ are both binomial, independent, with the same first parameter $p$ and second parameter $n_1$ and $n_2$, the sum $X + Y \sim b(p, n_1 + n_2)$. Why?

**Problems.**

1. Consider a binary random variable $b(p)$. Compute its variance as a function of $p$ and plot it. Where does it reach the maximum? What is the value of the maximum?

2. Compute the mean and variance of a uniform, $U(0,1)$. Compare its mean and variance with that of a $b(p = 0.5)$. Which is largest? Does this agree with our intuition?

3. Let $X$ and $Y$ be independent random variables with the following probability distributions:

| | Values of $x, y$ | | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | 3 |
| $P(x)$ | 0.1 | 0.3 | 0.4 | 0.2 |
| $P(y)$ | 0.5 | 0.3 | 0.1 | 0.1 |

   Using the characteristic (or moment generating) function, find the distribution of $Z = X + 2Y$

4. Let $X_n \sim b(p = \frac{1}{n})$, for $n = 1, 2, \ldots$. What does it converge to in distribution? What does it converge to in probability?

5. Let $X_n \sim b(p = \frac{1}{2})$, for $n = 1, 2, \ldots$ . What does it converge to in distribution? What does it converge to in probability?

6. When we have a batch of $N$ parts ($N$ very large) of which $d \ll N$ are defective and we pick three at random, will the number of defectives we find among these three be distributed as $b(p = \frac{d}{N}, n = 3)$? Why or why not? When, and when not, will that be a fairly good approximation?

7. In the case above, what is the exact probability of finding $k$ defectives when sampling $n$ units at random with and without replacement?

**Reading.** For the review probability items, any text you may have used in the previous course, or if you want something in English [2]. For the binary and binomial distribution, [1] § 7.3 and 7.4, or [3], Chapter 25.

# References


[1] J. Martín Pliego and L. Ruiz-Maya. *Estadística I : Probabilidad.* Ediciones AC, 2004. In the reserved collection, signature AL-519.2 MAR.

[2] Sheldon M. Ross. *A First Course In Probability.* Prentice Hall, 2003.

[3] A. Fz. Trocóniz. *Probabilidades. Estadística. Muestreo.* Tebar-Flores, Madrid, 1987.

**Soluciones**

1. La varianza es $pq = p(1-p)$. Derivando respecto a $p$ e igualando a cero tenemos $1 - 2p = 0$ de donde el valor que maximiza es $p = 0.5$. El valor máximo es $0.5 \times 0.5 = 1/4$. Es útil representar la función $p(1-p)$ (una parábola invertida) y constatar que el máximo está en $p = 0.5$ y los mínimos en $p = 0$ y $p = 1$.

2. La media, trivialmente, es 0.5, a medio camino entre 0 y 1. La varianza es,

$$\int_0^1 (x - 0.5)^2 dx = 1/12$$

Comparando con una $b(p = 0.5)$ vemos que la media es igual y la varianza 1/12 menos que 1/4. Esto, efectivamente, coincide con lo que esperaríamos: la uniforme toma valores entre los extremos 0 y 1, mientras que la binaria toma valores siempre en un extremo, 0 ó 1.

3. Tenemos, utilizando la función generatriz de momentos,

$$\begin{aligned} \varphi_X(u) &= 0.1 + 0.3e^u + 0.4e^{2u} + 0.2e^{3u} \\ \varphi_Y(u) &= 0.5 + 0.3e^u + 0.1e^{2u} + 0.1e^{3u} \\ \varphi_{2Y}(u) &= 0.5 + 0.3e^{2u} + 0.1e^{4u} + 0.1e^{6u} \end{aligned}$$

Por tanto, en virtud de la independencia,

$$\begin{aligned} \varphi_{X+2Y}(u) &= (0.1 + 0.3e^u + 0.4e^{2u} + 0.2e^{3u}) \times \\ &\qquad (0.5 + 0.3e^{2u} + 0.1e^{4u} + 0.1e^{6u}) \\ &= 0.1 \times 0.5 + 0.3 \times 0.5e^u + \ldots \end{aligned}$$

Los coeficientes de cada término $e^k$ dan la probabilidad de que $X + 2Y$ tome el valor $k$ y especifican completamente su distribución.

4. En distribución, converge a una causal que da probabilidad 1 al punto cero. En probabilidad converge al punto 0.

5. En distribución, converge a la distribución $b(p = 0.5)$ (¡todos los términos de la sucesión tienen la misma distribución, que es por tanto el límite!). En probabilidad, no converge.

6. No, no exactamente. Al retirar una pieza defectuosa del lote, el número de las que quedan es menor y por tanto la probabilidad de una nueva defectuosa ya no será $p = d/N$: no hay independencia entre extracciones. Sin embargo, si $d << N$ como se dice, la aproximación binomial será muy buena.

7. La distribución con reemplazamiento sería binomial. Sin reemplazamiento, sería hipergeométrica:

$$P(X = k) = \frac{\binom{d}{k}\binom{N-d}{n-k}}{\binom{N}{n}}.$$

**Pautas docentes.**

1. Introduciremos tanto la característica como la función generatriz de momentos, dejando claro que ambas pueden aparecer en problemas. En clase, podemos utilizar cualquiera.

2. Como analogía para mostrarles la utilidad de la función característica, suelen captar bien la siguiente: igual que una operación "complicada" como $x \times y$ se convierte en "simple" tras tomar logaritmos ($\log(xy) = \log(x) + \log(y)$), una operación funcional "complicada" como la convolución,

$$P(X+Y=z) = \sum_x P_X(x)P_Y(z-x)$$

se convierte en "simple" al tomar funciones características: $\psi_{X+Y}(u) = \psi_X(u)\psi_Y(u)$. Para que podamos sacar partido de esto es necesario, sin embargo, que sepamos recuperar $P_{X+Y}(z)$ de $\psi_{X+Y}(u)$ (al modo como recuperamos $xy$ tomando anti-logaritmo de $\log(xy)$).

3. A los que vengan de alguna carrera técnica puede ayudarles a situarse el hacerles notar que

$$\psi_X(u) = \int e^{iux} f_X(x)dx$$

no es más que una transformada de Fourier con otro nombre.

4. Puede ayudar a hacerles tragar la función característica (con su siempre repelente —para ellos— $i$) el hacerles notar que

$$\psi_X(u) = \int e^{iux} f_X(x)dx \leq \int |e^{iux}| f_X(x)dx = \int f_X(x)dx = 1$$

no tiene problemas de existencia, mientras que

$$\int e^{ux} f_X(x)dx$$

no está necesariamente acotada y puede diverger.

5. El ejercicio (7) da pie a mencionar e introducir la distribución hipergeométrica.