

Scientific journal
PHYSICAL AND MATHEMATICAL EDUCATION
Has been issued since 2013.

Науковий журнал
ФІЗИКО-МАТЕМАТИЧНА ОСВІТА
Видається з 2013.

ISSN 2413-158X (online)
ISSN 2413-1571 (print)



<http://fmo-journal.fizmatsspu.sumy.ua/>

Кобильник Т.П., Жидик В.Б. Методичні аспекти навчання дискретних випадкових величин з використанням статистичного середовища R. Фізико-математична освіта. 2018. Випуск 2(16). С. 58-62.

Kobylnyk T., Zhydyk V. Methodological Aspects Of Learning Discrete Random Variables With R Package. Physical and Mathematical Education. 2018. Issue 2(16). P. 58-62.

УДК 378.14

Т.П. Кобильник¹, В.Б. Жидик²

Дрогобицький державний педагогічний університет імені Івана Франка, Україна

¹kobylnyktaras@gmail.com, ²zhvb63@gmail.com

DOI 10.31110/2413-1571-2018-016-2-011

МЕТОДИЧНІ АСПЕКТИ НАВЧАННЯ ДИСКРЕТНИХ ВИПАДКОВИХ ВЕЛИЧИН З ВИКОРИСТАННЯМ СТАТИСТИЧНОГО СЕРЕДОВИЩА R

Анотація. Стаття присвячена методичним аспектам навчання студентів теорії ймовірностей, зокрема дискретних випадкових величин. Нами проаналізовано наукові та методичні джерела, порівняння, узагальнення з метою визначення стану проблеми та перспективних напрямків дослідження. Рекомендується використовувати статистичне середовище R як засіб навчання теорії ймовірностей та математичної статистики. Перш за все R – мова високого рівня та середовище для аналізу статистичних даних та їх графічної візуалізації. R – вільне програмне середовище з відкритим кодом, що поширюється в рамках проекту GNU. Широке навчання теорії ймовірностей та математичної статистики на основі пакетів цього середовища і всесвітня підтримка науковим співтовариством обумовили те, що наведення скриптів R поступово стає загально визнаним «стандартом» як в журнальних публікаціях, так і в неформальному спілкуванні науковців усього світу. Слід зазначити, що є кілька графічних інтерфейсів користувача для R (RStudio, R Commander та інші), але вони не такі хороші, як інтерфейсу Statistica або IBM SPSS.

Методи теорії ймовірностей використовуються в різних галузях науки. Здатність і вміння використовувати різні ймовірнісно-статистичні методи для аналізу експериментальних даних є однією з важливих умов для навчання вчителів інформатики. Впровадження ймовірнісно-статистичних методів аналізу в навчальний процес дозволяє підвищити рівень підготовки вчителів. Теорія ймовірностей є основою математичної статистики. У статті наведено переваги та недоліки R, приклади його використання для розв'язування деяких задач теорії ймовірностей, зокрема вивчення дискретних випадкових величин з біноміальними, геометричними, гіпергеометричними та розподілами Пуассона. Для деяких випадкових величин ми показали, як обчислити математичне сподівання та дисперсію, використовуючи пакет R. Подальші дослідження ми зосередимо на аналізі можливостей використання пакета R для статистичного моделювання, інтелектуального аналізу даних та методиці їх навчання в педагогічному університеті.

Ключові слова: теорія ймовірностей, дискретні випадкові величини, статистичне середовище R.

Постановка проблеми. Теорія ймовірностей є основою математичної статистики. Теорія ймовірностей та математична статистика є важливою складовою фундаментальної підготовки як фахівців з інформаційних технологій, так і педагогів. Для майбутніх вчителів (не тільки інформатики, фізики чи математики) знання методів теорії ймовірностей та математичної статистики є необхідними для здійснення статистичного аналізу даних у психолого-педагогічних дослідженнях. Тут слід відзначити навчальний посібник [1], у якому стисло подано основи теорії ймовірностей та математичної статистики та детально розглянуто методи перевірки статистичних гіпотез, що виникають під час опрацювання результатів психолого-педагогічних експериментів. Сьогодні неможливо уявити навчальний процес в університеті, в тому числі і педагогічному, без використання інформаційних технологій. У навчанні студентів інформатичних, фізико-математичних, природничих спеціальностей педагогічного університету використовується програмне забезпечення як загального, так і спеціального призначення. Одним з основних понять теорії ймовірностей є поняття випадкової величини.

Аналіз актуальних досліджень. Для навчання дисципліни «Теорія ймовірностей та математична статистика», як правило, обирають програму MS Excel [7]. У підручник [4], де подано основи теорії ймовірностей та математичної статистики для побудови графічних зображень, обчислення значень виразів, визначених інтегралів, аналізу статистичних даних, визначення числових характеристик розподілів ймовірностей, в тому числі статистичних, передбачається використання програми Graph [3]. Рідше обирають певну систему комп'ютерної математики, зокрема у посібнику [1] у

розділі, присвяченому вивченню випадкових величин, використовується система комп'ютерної математики Maple, що звільняє користувача від значної частини технічної роботи при обчисленнях. У працях [5, 8, 9, 10] проаналізовано розв'язування окремих задач з теорії ймовірностей та математичної статистики з використанням статистичного середовища R.

Мета статті: методичні аспекти використання статистичного середовища R у процесі навчання дискретних випадкових величин.

Методи дослідження. Для дослідження використовувались такі методи: системний науково-методологічний аналіз підручників і навчальних посібників, монографій, дисертаційних досліджень, статей і матеріалів науково-методичних конференцій з проблеми дослідження; спостереження навчального процесу; аналіз результатів навчання студентів у відповідності до проблеми дослідження; синтез, порівняння та узагальнення теоретичних положень, розкритих у науковій та навчальній літературі; узагальнення власного педагогічного досвіду та досвіду колег з інших закладів вищої освіти.

Виклад основного матеріалу. Для супроводу навчання дисципліни «Теорії ймовірностей та математична статистика» пропонується використовувати статистичне середовище R. Серед його переваг слід відзначити такі:

- є вільнопоширюваним програмним забезпеченням;
- існують реалізації під операційні системи Windows, Mac OS X, Linux;
- вбудована система допомоги та підказок;
- хороша графічна візуалізація подання даних та результатів їхнього аналізу;
- можливість самостійного написання необхідних функцій;
- багато літератури з R у вільному доступі.

Серед недоліків необхідно відзначити: на відміну від більшості комерційних програм R має не графічний інтерфейс, а інтерфейс командного рядка (орієнтація на програмування), тому треба знати необхідні для роботи функції та синтаксис мови програмування (цей недолік нівелюється наявністю графічних інтерфейсів);

Одним з основних понять теорії ймовірностей є поняття випадкової величини. Випадкові величини – це математичні моделі величин, що з'являються при проведенні стохастичних експериментів. Наприклад, виграш в азартних іграх, число відбувань події A при n незалежних випробувань, тривалість роботи пристроїв, помилки при вимірюваннях, екзаменаційні оцінки тощо. Випадкові величини можуть бути неперервними та дискретними. Випадкові величини мають деякий розподіл ймовірностей.

Для кожного з розподілів ймовірностей у статистичному середовищі R реалізовано чотири функції:

- 1) **dname** – ймовірність набуття випадковою величиною певного конкретного значення (для ДВВ) або значення функції щільності розподілу (для НВВ) – префікс d перед назвою розподілу;
- 2) **pname** – функція розподілу – префікс p перед назвою розподілу;
- 3) **qname** – квантілі розподілу – префікс q перед назвою розподілу;
- 4) **rname** – генерування значень випадкової величини за заданим розподілом – префікс r перед назвою розподілу.

Зауважимо, що **name** – це назва одного з розподілів ймовірностей на множині значень випадкової величини. У статті розглянуто такі розподіли ймовірностей дискретних випадкових величин як біноміальний (binom), Пуассона (pois), геометричний (geom) та гіпергеометричний (hyper).

Як правило, зміст навчання розділу теми «Дискретні випадкові величини» містить питання, пов'язані з вивченням розподілів ймовірностей, зокрема рівномірного, біноміального, Пуассона, геометричного, гіпергеометричного. У статистичному середовищі R серед інших реалізовані й наведені розподіли ймовірностей дискретних випадкових величин. Розглянемо кілька задач та проаналізуємо можливості використання статистичного середовища R для їх розв'язування.

Біноміальний розподіл ймовірностей. Дискретна випадкова величина ξ має біноміальний розподіл ймовірностей, якщо ймовірність її можливих значень обчислюється за формулою Бернуллі:

$$P_k = P(\{\xi = k\}) = C_n^k p^k \cdot q^{n-k}, k \in \overline{0,1, \dots, n},$$

де p – ймовірність «успіху», де q – ймовірність «невдачі» в одному випробуванні, причому $p + q = 1$.

Задача 1. Виконується 3 незалежних постріли по мішені. Ймовірність влучення при одному пострілі дорівнює 0,35. Побудувати ряд розподілу ймовірностей випадкової величини ξ – кількість влучень. Побудувати многокутник розподілу ймовірностей випадкової величини ξ . Обчислити ймовірність $P(\xi > 1)$. Знайти математичне сподівання та дисперсію випадкової величини ξ .

Розв'язування. Очевидно, що множиною можливих значень випадкової величини ξ є множина $\Omega = \{0,1,2,3\}$. Згідно умови задачі ймовірність «успіху» $p = 0.35$, ймовірність «невдачі» $q = 1 - p = 1 - 0.35 = 0.65$. За формулою Бернуллі

$$P_k = P(\{\xi = k\}) = C_3^k p^k \cdot q^{3-k}, k \in \Omega.$$

Надаючи k значень 0,1,2,3, отримаємо ряд розподілу ймовірностей (див.таблиця 1).

Таблиця 1.

ξ	0	1	2	3
$P_k = P(\{\xi = k\})$	0.274625	0.443625	0.238875	0.042875

$$P(\xi > 1) = P(\xi = 1) + P(\xi = 2) = 0.238875 + 0.042875 = 0.28175.$$

Згідно формул для обчислення математичного сподівання, дисперсії для випадкової величини, розподіленої за біноміальним розподілом, отримується:

$$M(\xi) = np = 3 \cdot 0.35 = 1.05; D(\xi) = npq = 3 \cdot 0.35 \cdot 0.65 = 0.6825.$$

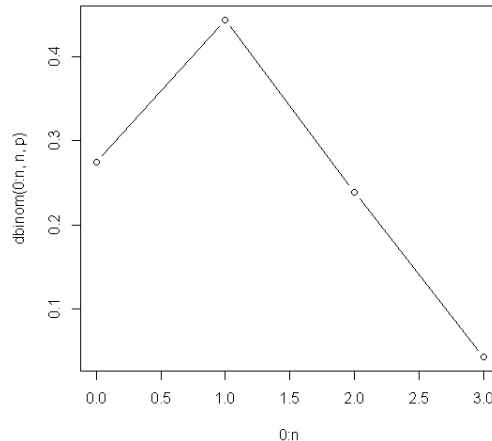
Реалізація в R.

- 1) побудова ряду розподілу ймовірностей випадкової величини ξ :
- ```
> n=3;p=0.35;A <- data.frame(P = dbinom(0:n, n, p))
> rownames(A) <- 0:n
> A
```

```
 P
0 0.274625
1 0.443625
2 0.238875
3 0.042875
```

- 2) за функцією plot будемо многокутник розподілу ймовірностей випадкової величини  $\xi$ :

```
> plot(0:n,dbinom(0:n,n,p),type="b")
```



- 3) обчислення ймовірності  $P(\xi > 1)$  можна таким чином:

а) використаємо функцію sum:

```
> #обчислення ймовірності більше одного влучення
> sum(dbinom(2:n,n,p))
[1] 0.28175
```

б) використаємо функцію pbinom:

```
> pbinom(1,n,p,lower=FALSE)
[1] 0.28175
```

*Зауваження 1.* У функції pbinom використано значення lower=FALSE. Таким чином обчислюється ймовірність  $P(\xi > x)$ . За замовчуванням є значення lower=TRUE, тобто обчислюється ймовірність  $P(\xi \leq x)$ .

*Зауваження 2.* У вітчизняній літературі з теорії ймовірностей та математичної статистики, зокрема [2, 5], функція розподілу  $F(x) = P(\xi < x)$ . У зарубіжній літературі, зокрема [7, 8, 9],  $F(x) = P(\xi \leq x)$ , тобто права межа інтервалу включається. Саме так, як у зарубіжній літературі з теорії ймовірностей та математичної статистики, у програмному середовищі R реалізована функція розподілу ймовірностей, на що треба звертати увагу під час розв'язування задач.

- 4) для знаходження математичного сподівання  $M(\xi)$  та дисперсії  $D(\xi)$  випадкової величини  $\xi$  зручно використовувати відповідні функції wtd.mean та wtd.var з пакету Hmisc:

```
> #математичне сподівання
> wtd.mean(0:n,pp)
[1] 1.05
> #дисперсія
> wtd.var(0:n,pp,method="ML")
[1] 0.6825
```

**Розподіл Пуассона.** Дискретна випадкова величина  $\xi$  має розподіл Пуассона, якщо ймовірності її можливих значень обчислюється за формулою

$$P_k = P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda},$$

де  $k = 0, 1, 2, \dots, n$ ,  $\lambda = np$ . Розподіл Пуассона пов'язаний з біноміальним таким чином. При необмеженому збільшенні  $n$  за умов  $n \rightarrow \infty$ ,  $p \rightarrow 0$  (або  $q \rightarrow 0$ ) і  $np = \lambda = \text{const}$  біноміальний розподіл імовірностей асимптотично наближається до розподілу Пуассона.

*Задача 2.* Визначити ймовірність того, що на АЗС знаходиться один або принаймні один автомобіль, якщо середнє число автомобілів, які знаходяться у даний термін часу на АЗС дорівнює 4.

*Розв'язування.* За умовою задачі маємо  $\lambda = 4$ . Тоді за формулою Пуассона ймовірність того, що на АЗС знаходиться принаймні один автомобіль:

$$P(\xi = 1) = \frac{\lambda^1}{1!} e^{-\lambda} = 4 \cdot e^{-4} \approx 0.07326.$$

Ймовірність того, що на АЗС знаходиться більше одного автомобіля:

$$P(\xi > 1) = 1 - P(\xi = 0) = 1 - \frac{\lambda^0}{0!} e^{-\lambda} = 1 - e^{-4} \approx 0.981684.$$

*Реалізація в R.*

```
> dpois(1,4)#ймовірність заходження 1 автомобіля
[1] 0.07326256
> 1-dpois(0,4)#ймовірність знаходження більше одного автомобіля
[1] 0.9816844
```

**Геометричний розподіл.** Випадкова величина  $\xi$  має геометричний розподіл з параметром  $p$  ( $p > 0$ ), якщо  $P_{\xi}(n) = (1 - p)^{n-1}p, n \in \overline{1,2,..}$

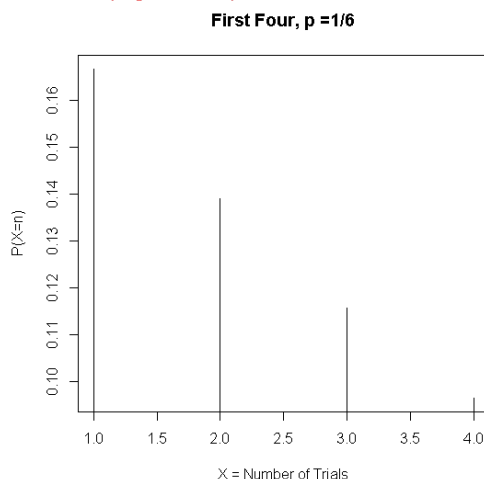
**Задача 3.** Гральний кубик підкидається до першої появи цифри 3. Знайти ймовірність того, що гравцю доведеться 4 рази підкидати кубик.

*Розв'язування.* За умовою задачі  $n = 4, p = 1/6$ . Тоді згідно формули

Згідно формули,  $P_{\xi}(4) = \left(1 - \frac{1}{6}\right)^{4-1} \frac{1}{6} = \frac{125}{1296} \approx 0.09640506$ .

*Реалізація в R.*

```
> n=4;p=1/6;dgeom(n-1,1/6)
[1] 0.09645062
> n=1:4;p=1/6;
> plot(n,dgeom(n-1,p),xlab = "X = Number of Trials", ylab = "P(X=n)",
+ type = "h", main = "First Four, p =1/6")
```



**Гіпергеометричний розподіл.** Випадкова величина  $\xi$  має гіпергеометричний розподіл, з параметрами  $(N, M, n), n \leq M, n \leq N - M$ , якщо  $P(\xi = m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}$ , де  $m = 0, 1, 2, \dots, n$ .

Нехай в урні міститься  $N$  куль, серед них  $M$  – білі, решта  $(N - M)$  – чорні. Кількість білих куль серед  $n$  ( $n \leq M, n \leq N - M$ ) навмання вибраних з урни має гіпергеометричний розподіл з параметрами  $(N, M, n)$ .

**Задача 4.** В урні міститься сім однакових за розміром куль, серед яких три білих, решта – чорні. Побудувати ряд розподілу ймовірностей випадкової величини  $\xi$  – кількості куль білого кольору серед 4 вибраних. Обчислити математичне сподівання  $M(\xi)$  та дисперсію  $D(\xi)$ .

*Розв'язування.* Випадкова величина  $\xi$  має гіпергеометричний розподіл, згідно якого ймовірності для кожного можливого значення випадкової величини обчислюється за формулою  $P(\xi = k) = \frac{C_3^k C_4^{4-k}}{C_7^4}, k \in \overline{0,3}$ . Обчисливши ймовірності, отримаємо такий ряд розподілу (таблиця 2).

**Таблиця 2.**

| $\xi$                  | 0              | 1               | 2               | 3              |
|------------------------|----------------|-----------------|-----------------|----------------|
| $P_k = P(\{\xi = k\})$ | $\frac{1}{35}$ | $\frac{12}{35}$ | $\frac{18}{35}$ | $\frac{4}{35}$ |

*Реалізація в R.*

```
> p=dhyper(0:3,3,4,4);round(cbind(0:3,p),4)
 p
[1,] 0 0.0286
[2,] 1 0.3429
[3,] 2 0.5143
[4,] 3 0.1143
```

Слід звернути увагу на те, що тип чисел у статистичному середовищі R отримується у дійсному форматі, а не у вигляді звичайного дробу.

**Висновки.** У статті проаналізовано можливості використання статистичного середовища R для розв'язування задач теорії ймовірностей, зокрема з розділу дискретні випадкові величини. Статистичне середовище R є однією з кращих програм для навчання студентів інформатичних та фізико-математичних спеціальностей статистичних методів аналізу експериментальних даних, враховуючи безкоштовність та можливість роботи під управлінням різних операційних систем, в тому числі і Linux. Серед недоліків R слід зазначити орієнтацію на програмування (хоча це може нівелюватися наявністю різноманітних графічних інтерфейсів). Подальші дослідження будуть спрямовані на аналіз можливостей використання статистичного середовища R для розв'язування задач статистичного моделювання, інтелектуального аналізу даних та методики їх навчання у педагогічному університеті.

## Список використаних джерел

1. Бабенко В.В. Основы теории вероятностей и статистические методы анализа данных в психологических и педагогических экспериментах: навч. посібник. Львів: Видавничий центр ЛНУ імені Івана Франка, 2009. 184 с.
2. Гихман И.И., Скороход А.В., Ядренко М.И. Теория вероятностей и математическая статистика. К.: Вища школа, 1979. 408 с.
3. Жалдак М.І., Горошко Ю.В., Вінниченко Є.Ф. Математика з комп'ютером : посіб. для вчителів. 2-ге вид. К. : НПУ ім. Драгоманова, 2009. 282 с.
4. Жалдак М.І., Кузьміна Н.М., Михалін Г.О. Теорія ймовірностей і математична статистика: Підручник для студентів фізико-математичних та інформатичних спеціальностей педагогічних університетів. Видання третє, перероблене і доповнене. Київ: НПУ імені М.П. Драгоманова, 2015. 705 с.
5. Зарядов И.С. Статистический пакет R: теория вероятностей и математическая статистика: Учебно-методическое пособие. М.: РУДН, 2010. 141 с.
6. Колмогоров А.Н. Основные понятия теории вероятностей. М.: Наука, 1974. 132 с.
7. Руденко В. М. Математична статистика. Навч. посіб. К.: Центр учбової літератури, 2012. 304 с.
8. Dekking F.M., Kraaikamp C., Lopuhaä H.P., Meester L.E. A Modern Introduction to Probability and Statistics: Understanding Why and How. Springer-Verlag London Limited, 2005. 483 p.
9. Durrett R. Probability: Theory and Examples. Fourth Edition. Cambridge: Cambridge University Press, 2010. 440 p.
10. Ugarte M.D., Militino A.F., Arnholt A.T. Probability and statistics with R. Boca Raton, London, New York: CRC Press, Taylor&Francis Group, 2008. 700 p.

## References

1. Babenko V.V. Osnovy teorii ymovirnostei i statystychni metody analizu danykh u psykholohichnykh i pedahohichnykh eksperymentakh: navch. posibnyk. Lviv: Vydavnychiy tsentr LNU imeni Ivana Franka, 2009. 184 s. (in Ukrainian)
2. Gihman I.I., Skorohod A.V., Jadrenko M.I. Teorija verojatnostej i matematicheskaja statistika. K.: Vishha shkola, 1979. 408 s. (in Russian)
3. Zhaldak M.I., Horoshko Yu.V., Vinnychenko Ye.F. Matematyka z kompiuterom : posib. dlia vchyteliv. 2-he vyd. K. : NPU im. Drahomanova, 2009. 282 s. (in Ukrainian)
4. Zhaldak M.I., Kuzmina N.M., Mykhalin H.O. Teoriia ymovirnostei i matematychna statystyka: Pidruchnyk dlia studentiv fizyko-matematychnykh ta informatychnykh spetsialnostei pedahohichnykh universytetiv. Vydannia tretie, pereroblene i dopovnene. Kyiv: NPU imeni M.P. Drahomanova, 2015. 705 s. (in Ukrainian)
5. Zarjadov I.S. Statisticheskij paket R: teorija verojatnostej i matematicheskaja statistika: Uchebno-metodicheskoe posobie. M.: RUDN, 2010. 141 s. (in Russian)
6. Kolmogorov A.N. Osnovnye ponjatija teorii verojatnostej. M.: Nauka, 1974. 132 s. (in Russian)
7. Rudenko V. M. Matematychna statystyka. Navch. posib. K.: Tsentr uchbovoi literatury, 2012. 304 s. (in Ukrainian)
8. Dekking F.M., Kraaikamp C., Lopuhaä H.P., Meester L.E. A Modern Introduction to Probability and Statistics : Understanding Why and How. Springer-Verlag London Limited, 2005. 483 p. (in English)
9. Durrett R. Probability: Theory and Examples. Fourth Edition. Cambridge: Cambridge University Press, 2010. 440 p. (in English)
10. Ugarte M.D., Militino A.F., Arnholt A.T. Probability and statistics with R. Boca Raton, London, New York: CRC Press, Taylor&Francis Group, 2008. 700 p. (in English)

## METHODOLOGICAL ASPECTS OF LEARNING DISCRETE RANDOM VARIABLES WITH R PACKAGE

Taras Kobylnyk, Volodymyr Zhydyk

Drohobych Ivan Franko State Pedagogical University, Ukraine

**Abstract.** The article is devoted to methodological aspects of teaching students of probability theory, in particular discrete random variables. Scientific and methodological sources were analyzed, comparison, generalization was used to determine the state of the problem and perspective directions of its solution. R is recommended to be used as a learning tool for probability theory and mathematical statistics. First of all R is a high-level language and an environment for data analysis and graphics. It is a GNU project. It is a free software environment with open source. Learning of statistical analyses with R and global scientific community support has resulted that R scripts are the standard for both of articles and in informal communication of scientists around the world. It should be noted there are several Graphical User Interfaces for R (RStudio, R Commander and others), but they are not as good as the interfaces Statistica or IBM SPSS.

Probability theory methods is being used in various fields of science. The ability and skills to use various statistical methods for analyzing experimental data is one of the important conditions for the training of computer science teachers. The introduction of statistical methods of analysis in the educational process makes it possible to improve the level of teacher training. The probability theory is the basis of mathematical statistics. In the article we have presented the advantages and disadvantages of R, examples of its use for solving some problems of probability theory, in particular to study discrete random variables with a binomial, geometric, hypergeometric, and Poisson distributions. For some random variables, we showed how to calculate the mathematical expectation and variance using the R package. Further research we will focus on the analysis of the possibilities of using the R package for solving problems of statistical modeling, data mining and teaching methodology at the pedagogical university.

**Keywords:** Probability Theory, Discrete Random Variables, R package.