TP 2: variables continues

anna.melnykova@univ-avignon.fr

Le but de cette partie c'est de continuer explorer les possibilités du logiciel R et commencer de travailler avec les variables continues. Plus précisément, on vise à assimiler: *simulation de variables* continues par l'inversion, *outils graphiques:* histogrammes, graphiques de densité, *statistiques descriptives*: moyenne, médiane, quantiles

Commandes utiles:

- `runif(n)` — créer un vecteur qui contient n réalisations d'une variable uniforme $\mathcal{U}[0,1]$, `rexp(n, rate = r)` — d'une variable exponentielle avec le paramètre $\lambda = r$ (par defaut $\lambda = 1$)
- `hist(x)` — construire une histogramme pour visualiser le vecteur `x`.
- `dexp(x), pexp(x)` — calculer la densité dans le point x et la probabilité $\mathcal{P}(X < x)$ pour une variable exponentielle standard (avec $\lambda = 1$).

Exercice 1: Loi exponentielle

On suppose que la durée d'attente (en minutes) de la première alerte après 22 heures dans un service de secours suit à peu près une loi exponentielle de paramètre $\lambda = 0.1$. On cherche à étudier la distribution de la variable aléatoire qui modélise le temps d'attente.

Rappelons-nous que la densité et la fonction de répartition pour la loi exponentielle sont données par:

$$f(x) = \lambda e^{-\lambda x}, \quad F(x) = 1 - e^{-\lambda x}$$

Une des méthodes de simulation de variables aléatoires consiste en l'inversion de la fonction de répartition. L'idée est simple: on simule les variables uniformes $U \sim \mathcal{U}([0,1])$, ou chaque observation u_i nous donne la probabilité $P(X \leq x_i) = F(x_i)$ d'observer x_i, pas encore connu. Puis on calcule x_i en inversant $F(x_i)$. Autrement dit, on fait deux étapes:

1. On tire $u \sim \mathcal{U}([0,1])$
2. On calcule $x = F^{-1}(u)$

- Tout d'abord, calculez (mathématiquement) la fonction inverse de $F(x)$. Pour rappel, pour trouver une fonction inverse, il faut résoudre une équation $y = F(x)$ en fonction de y (i.e., trouver une expression $x = F^{-1}(y)$).

- Maintenant, on cherche à simuler les observations issues de la loi exponentielle en utilisant la loi uniforme. Puis, on superpose la densité théorique de la loi exponentielle avec $\lambda = 0.1$ pour vérifier que l'échantillon obtenu correspond bien à la loi théorique.

```
lmb = 0.1
n = 1000
U_n = runif(n)
X_n # remplacer par votre fonction inverse
hist(X_n, prob = T)
curve(dexp(x, rate = lmb), col = "red", add = TRUE)
```

- Maintenant, on va calculer la probabilité pour qu'il n'y a ait aucune alerte avant 22h30:

```
mean(X_n>30) # Probabilité empirique
pexp(30, lmb, lower.tail = FALSE) # Probabilité théorique
```

- Estimer empiriquement la probabilité pour qu'il n'y a ait aucune alerte avant 23h.

- Posez $n = 100000$ et reéxecutez les commandes précédents
- Evaluez empiriquement la probabilité pour qu'il n'y a ait aucune alerte avant 23h sachant qu'il n'y a aucune alerte avant 22h30? *Indication*: première commande crée un vecteur qui contient que les alertes qui se passent après 22h30. Utilisez-le pour donner l'approximation numérique de la probabilité $P(X > 60|X > 30)$.

```
X_n_30 <- X_n[which(X_n>30)]
mean(X_n_30>60) # Probabilité empirique
pexp(30, lmb, lower.tail = FALSE) # Probabilité théorique
```

- Est-ce que les valeurs numériques correspondent aux valeurs exactes, calculées en TD?

Exercice 2: Loi normale

Rappelons-nous que la loi normale $\mathcal{N}(\mu, \sigma^2)$ est définie par sa fonction de densité

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Si $X \sim \mathcal{N}(\mu_X, \sigma^2)$ et $Y \sim \mathcal{N}(\mu_Y, \sigma^2)$, alors $X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma^2)$, et $\forall a \in \mathbf{R}$ $aX \sim \mathcal{N}(a\mu_X, a^2\sigma^2)$.

Soit X une variable aléatoire de loi normale $\mathcal{N}(0, 1)$.

1. Exécutez la commande `?pnorm`. Quelle est la différence entre les fonctions `pnorm`, `qnorm`, `rnorm` et `dnorm`?

2. Déterminer $P(X \leq 0.41)$, $P(X < -0.2)$, $P(X \geq -1.54)$, $P(-0.63 \leq X \leq 1.2)$ en se servant de la commande `pnorm`.

3. Déterminer u tel que $P(X \leq u) = 0.75$ en se servant de la commande `qnorm`.

4. Simuler 100 observations de X en utilisant la commande $rnorm(n)$ et visualisez-les avec un histogramme.
5. Calculez l'approximation des probabilités obtenues dans la question 1. Donnez l'approximation de u dans question 2 en utilisant la commande `summary`.

6. Augmentez la taille d'échantillon et recalculez. Que observez-vous?