

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Statistika merupakan ilmu pengetahuan yang berhubungan dengan teknik pengumpulan, pengolahan, atau penganalisisan dan penarikan kesimpulan berdasarkan kumpulan data dan penganalisisan yang dilakukan (Sudjana, 1996). Perkembangan ilmu statistika, dapat dikatakan tidak pernah berhenti. Pembaharuan selalu ada untuk melengkapi teori-teori ataupun metode-metode yang telah ada. Hal itu ditujukan untuk memudahkan dalam pemecahan setiap masalah yang ada.

Hasil dari percobaan atau pengolahan data digambarkan sebagai nilai-nilai numerik yang disebut dengan variabel acak. Ada dua jenis variabel acak, yaitu variabel acak diskrit dan variabel acak kontinu. Jika himpunan hasil dari variabel acak berhingga atau dapat dihitung, maka variabel tersebut disebut sebagai variabel acak diskrit. Sedangkan jika semua hasil yang mungkin dari variabel acak merupakan nilai dalam suatu interval, maka disebut sebagai variabel acak kontinu (Bain & Engelhardt, 1992).

Dalam statistika, dikenal juga distribusi peluang. Distribusi peluang adalah fungsi matematika yang menggambarkan kemungkinan terjadinya berbagai kejadian setelah melakukan percobaan. Distribusi peluang dipergunakan untuk menentukan bagaimana nilai peluang didistribusikan pada data (Otaya, 2016).

Berdasarkan jenis variabel acaknya, distribusi peluang dibagi menjadi distribusi peluang diskrit dan distribusi peluang kontinu. Yang termasuk distribusi peluang diskrit, diantaranya distribusi Poisson, distribusi Binomial, distribusi Geometrik, dan distribusi Bernoulli. Sementara itu, yang termasuk dalam distribusi peluang kontinu diantaranya distribusi Normal, distribusi Gamma, distribusi Eksponensial, distribusi Beta, dan distribusi Pareto. Dari berbagai distribusi peluang kontinu yang ada, distribusi normal merupakan distribusi yang paling banyak digunakan. Hal tersebut dikarenakan parameter dari distribusi normal yang langsung menyatakan mean dan varians dari data, sehingga dapat

mempermudah dalam analisis statistik. Akan tetapi, distribusi normal hanya cocok digunakan untuk data yang mempunyai kurvanya yang simetris.

Distribusi Pareto yang merupakan distribusi peluang kontinu adalah distribusi yang pertama kali ditemukan oleh seorang ekonom dan sosialis Italia bernama Vilfredo Pareto. Pareto menciptakan formula matematika di awal abad ke-20 yang menggambarkan ketidaksetaraan dalam distribusi kekayaan yang ada di negara asalnya, yaitu Italia. Distribusi Pareto adalah salah satu bagian dari keluarga distribusi *power law*, dimana dalam distribusi ini, rata-rata atau standar deviasinya akan dipengaruhi oleh nilai pangkat fungsi kepadatan peluangnya. Distribusi Pareto memiliki dua parameter yang biasa disebut parameter skala dan parameter bentuk. Distribusi Pareto umumnya digunakan dalam bidang sosial, ekonomi, bisnis, asuransi, maupun politik (Arnold, 2004).

Beberapa penelitian yang mengkaji perkembangan dari distribusi Pareto, diantaranya adalah *Estimation In The Pareto Distribution* oleh Mette Rytgaard yang mengkaji mengenai estimasi parameter dari distribusi Pareto, *An Extended Pareto Distribution* oleh M. E. Mead, yang mengkaji mengenai kecocokan distribusi Pareto untuk digunakan dalam data uji hidup, dan *Weibull-Pareto Distribution and Its Applications* yang mengkaji mengenai karakteristik yang dimiliki oleh distribusi Weibull-Pareto, pengestimasian parameter, dan pengaplikasiannya (Alzaatreh, Lee, & Famoye, 2013). Dari penelitian terdahulu, diketahui bahwa distribusi Pareto dikategorikan sebagai distribusi *heavy tailed*.

Distribusi *heavy tailed* adalah distribusi yang mampu menggambarkan data yang memiliki nilai peluangnya besar ketika nilai pengamatan dari data itu bernilai besar. Sementara distribusi *light tailed* adalah distribusi yang mampu menggambarkan data yang memiliki nilai peluang kecil saat nilai pengamatan dari data itu bernilai besar. Adapun salah satu distribusi yang biasa digunakan untuk menggambarkan data yang *heavy tailed* adalah distribusi Pareto. Akan tetapi, distribusi Pareto hanya mempunyai satu bentuk grafik dari fungsi kepadatan peluang yang berbentuk eksponensial. Sehingga karena hal tersebut, distribusi Pareto dianggap kurang fleksibel dalam menggambarkan data. Pada tahun 2015, Mahdavi dan Kundu mengusulkan suatu metode baru dengan memperkenalkan

suatu parameter tambahan untuk menambah fleksibilitas pada suatu distribusi. Metode yang diusulkan tersebut merupakan *Alpha Power Transformation* (APT).

APT merupakan salah satu metode yang sederhana dalam menambahkan parameter untuk menggeneralisasi distribusi yang sudah ada berdasarkan bentuk fungsi kepadatan peluang (*pdf*) dan fungsi distribusi kumulatif (*cdf*) nya. Selain itu, metode ini dapat meningkatkan fleksibilitas dari karakteristik distribusi yang sudah ada sebelumnya. Transformasi ini juga dapat digunakan untuk model dengan fungsi hazard yang monoton dan non-monoton, serta dapat memberikan fleksibilitas yang lebih baik dibandingkan dengan modifikasi model lainnya yang memiliki banyak parameter yang sama ataupun lebih (Elbatal & Al, 2018).

Pada penelitian skripsi ini, akan dibahas mengenai pengembangan dari distribusi Pareto dengan menambahkan transformasi *alpha power*, yaitu distribusi APT-Pareto. Distribusi APT-Pareto adalah modifikasi atau pengembangan dari distribusi Pareto dengan adanya tambahan parameter  $\alpha$  ke dalamnya. Sehingga, karakteristik yang dimilikinya pun berbeda dengan karakteristik yang dimiliki oleh distribusi Pareto. Berdasarkan hal tersebut, peneliti tertarik untuk membahas mengenai karakteristik yang dimiliki oleh distribusi APT-Pareto yang akan dituangkan dalam skripsi dengan judul “Distribusi APT-Pareto dan Aplikasinya”.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, maka penulis merumuskan beberapa masalah yang akan dibahas, yaitu:

1. Apa saja karakteristik yang dimiliki oleh distribusi APT-Pareto?
2. Bagaimana pengaplikasian distribusi APT-Pareto terhadap data waktu bertahan hidup pasien *myelogenous leukemia*?

## 1.3 Tujuan

Berdasarkan rumusan masalah yang telah diuraikan, maka tujuan dari penulisan skripsi ini yaitu sebagai berikut:

1. Untuk mempelajari tentang distribusi APT-Pareto dan mengetahui karakteristiknya.

2. Untuk mengetahui bagaimana hasil dari pengaplikasian distribusi APT-Pareto pada waktu bertahan hidup pasien *myelogenous leukemia*.

#### **1.4 Manfaat**

Berdasarkan rumusan masalah dan tujuan penulisan yang telah diuraikan di atas, maka manfaat yang penulis harapkan dalam penulisan skripsi ini yaitu sebagai berikut:

1. Manfaat Teoritis:
  - a. Penelitian ini dapat menjadi sumber ilmu pengetahuan dan memperluas wawasan mengenai perkembangan distribusi statistik dengan metode transformasi, khususnya mengenai distribusi APT-Pareto.
  - b. Penelitian ini dapat menjadi wadah untuk menerapkan ilmu yang telah didapat selama perkuliahan.
2. Manfaat Praktis  
Mengetahui hasil pengaplikasian distribusi APT-Pareto pada data waktu bertahan hidup pasien *myelogenous leukemia*.